10

Action-Oriented Models of Cognitive Processing

A Little Less Cogitation, A Little More Action Please

James Kilner, Bernhard Hommel, Moshe Bar,
Lawrence W. Barsalou, Karl J. Friston, Jürgen Jost,
Alexander Maye, Thomas Metzinger, Friedemann Pulvermüller,
Marti Sánchez-Fibla, John K. Tsotsos, and Gabriella Vigliocco

Abstract

This chapter considers action-oriented processing from a model-oriented standpoint. Possible relationships between action and cognition are reviewed in abstract or conceptual terms. We then turn to models of their interrelationships and role in mediating cognitively enriched behaviors. Examples of theories or models inspired by the action-oriented paradigm are briefly surveyed, with a particular focus on ideomotor theory and how it has developed over the past century. Formal versions of these theories are introduced, drawing on formulations in systems biology, information theory, and dynamical systems theory. An attempt is made to integrate these perspectives under the enactivist version of the Bayesian brain; namely, active inference. Implications of this formalism and, more generally, of action-oriented views of cognition are discussed, and open issues that may be usefully pursued from a formal perspective are highlighted.

Existing Schema for Action and Cognition

Before considering the form and consequences of models that take an action-oriented view of cognition, it is worth considering how the relationship between these two processes has been described. Here, we consider four schemata which capture different notions of how cognition and action could be coupled (Figure 10.1). It is important to stress that these schemata are not models but

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

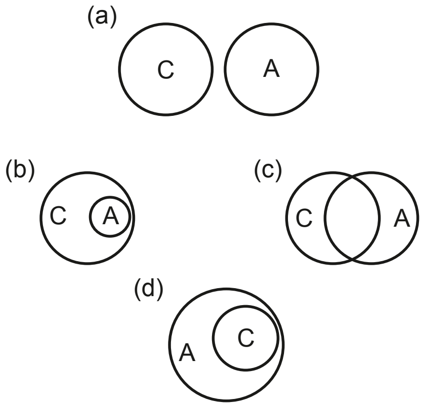


Figure 10.1 Schemata of action-cognition relationships.

rather depictions of different views on cognition and action. Crucially, differences in the schemata do not reflect a fundamental difference in the nature of the coupling but rather in how cognition and action are defined.

At one extreme, in Figure 10.1a, the two processes can be considered to be entirely independent processes: action is simply the behavioral output of the cognitive process (e.g., in the classical sandwich conception of the mind). In this open-loop formulation, action does not, and indeed cannot, influence cognition. Although not explicitly stated, the scheme depicted in Figure 10.1a is implicitly assumed in many models of high-level human cognitive functions used in behavioral and imaging experiments: input is carefully manipulated and output (action) is kept to a minimum (e.g., key pressing) in order to focus on internal processes. Here, action is considered to be a necessary output to disclose internal operations. For example, in a typical experiment that investigates language processing, individuals are presented with written or spoken words/sentences and asked to make a key press decision about them. Such studies, which still constitute the majority of cognitive science and neuroscience studies of high-level cognition, are completely silent as to whether action might play a role.

Alternative approaches (Figure 10.1b–d) consider how action and cognition depend on each other. In addition, action can be explicitly considered to be a subset of cognition or a largely overlapping process (Figure 10.1b and c). This commonly held view defines cognition as information processing à la Neisser. Accordingly, most, but not all, cognition relates to action: cognition influences action, and action influences cognition. This influence is typically, but not always, constructive: cognition and action can exist in harmony without collapsing into one another. They are heavily intertwined, but they are not the same, and none exists solely for the benefit of the other. Just like attention, working memory, and consciousness, cognition and action are intimately related, yet

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

independent. The final scheme (Figure 10.1d) represents a more "enactive" form: cognition is a subset of action and/or cognition subserves action.

In these schemata, action is often only considered as an external goal-directed movement (as opposed to an internal/mental action). Although actions are often defined as goal-directed movements, they could also be considered on a continuum between motor movements and goal-directed actions. For example, the babbling infant may produce the same syllable [ti:] as a one-year-old repeating a word and an older child asking for the beverage. In development, successively more distant goals and predictions (e.g., getting tea, ordering on behalf of someone in a restaurant to please that person) seem to be coupled with the movement representations to yield the cognitive action representation. Because of the relationship between their putative neuronal circuit representations, it may make sense to see movements and goal-directed actions on a continuum.

From a cognitive perspective, an action results from mediating processes that constitute proximal causes of the action relative to the distal stimulus and subsequent sensory processes. An action emanating from a cognitive system has an important conceptual interpretation above and beyond its motor features (often public or social); for example, donating to charity over the Internet. Such an action includes motoric features (reading a screen and typing on a keyboard), while invoking the notion of moving money from one's bank account to an organization to help relieve suffering in a distant group of people. These latter aspects play an important role in conceiving of the action in the first place: donating to charity would not be possible without understanding money, charities, donation, and suffering—and all the causal relations among them—and executing the resultant action effectively and monitoring whether the intended outcome occurred (see Barsalou, this volume). Arguably the most important actions that humans perform are ones that acquire physical resources (e.g., food, shelter, wealth), alter the physical environment (e.g., clearing land, farming), and develop and use technology (tools) to achieve goals as well as establish, maintain, and revise social relations and social status. Crucially, and especially as a result of language and communication, these actions call on cultural institutions, artifacts, and knowledge.

Models of Action-Oriented Processing

Very few models are actually based on a relationship between action and cognition, as illustrated in the third schemata (Figure 10.1c). One important example of an attempt to close the loop connecting brain and environment is the ideomotor theory, as attributed to Lotze (1852), Harless (1861), and James (1890). Its original formulation tried to explain how people acquire voluntary control of their actions, even though they do not seem to have privileged (conscious) access to the motor system. The idea is that agents start interacting with

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science," Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18, series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

their environment by motor babbling (executing random movements), sensing the reafferent information resulting from these movements (i.e., the self-produced changes in perception), and creating bidirectional associations between the neural pattern producing the movement and the neural pattern representing the reafferent information (Hebbian learning). Given that agents can activate the reafferent codes endogenously by imagining the respective events, these bidirectional associations provide them with retrieval cues to the associated motor patterns. In such a way, the movements can now be executed intentionally. The theory of event coding (TEC) has extended and generalized this approach in various ways (Hommel et al. 2001). First, it claims that perceptual codes and action patterns are represented in a distributed, nonsymbolic fashion. Second, it assumes that perceptual codes and action patterns are integrated into sensorimotor event files; that is, into networks of codes representing the perceptual and movement-related aspect of a given sensorimotor event. Third, it assumes that to represent a given event, the components of a given event file are weighted according to their relevance for the given action goal (intentional weighting; Memelink and Hommel 2013). For instance, when grasping an object, shape, and location, features will be weighted more strongly (and thus contribute more to the representation of the object-grasping event) than color features. Finally, TEC claims that perception and action are the same thing: the process of perceiving an event entails moving in ways that orient one's receptors toward the event of interest, so as to register its perceptual features, and the process of producing an event (i.e., acting) involves moving in ways to generate particular perceptual features which are then sensed and compared to the expected outcomes. That is, both perception and action actively generate reafferent input but the specificity to which this input is predicted is often lower for what we call perception than for what we call action. In essence, TEC assumes that the basis of human cognition is sensorimotor in nature (event files). However, it does not explicitly rule out the possibility of more abstract cognitive codes that are derived from event files.

Other models have been proposed that also make the link between action and cognition explicit. For example, models of sensorimotor representations of grasping movements in frontoparietal cortex can be used to explain the perception of actions as well as the "simulation" or "mentalizing" about actions (Arbib et al. 2000; Jeannerod et al. 1995). A model of frontotemporal cortex shows the emergence of linked action-perception mechanisms from sensorimotor information and functional implications of such learning for working memory (Pulvermüller and Garagnani 2014). Hebb-type learning leads to a strengthening of neuronal connections in a pool of sensorimotor neurons, which implies that activity will be maintained longer in the pool. These models show how higher cognitive functions (mentalizing, memory, and so on) can develop in specific neuroanatomical structures on the basis of associative learning of correlated motor and sensory activity. Thus they provide concrete implementations of the functional parallelism between cognition and action.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

Over and above accounting for the emergence of mechanisms for higher cognition, neuroanatomically grounded action-perception models may explain the specific and dissociable brain areas carrying particular cognitive functions.

Here we have provided a few examples of theoretical approaches that link (or unify) action and cognition in conceptual terms. In what follows, we revisit the same ideas from a formal perspective, trying to identify the decomposition of states and their dynamics that constitute the problem at hand. In particular, we consider the optimality principles inherent in ideomotor theory and related developments.

Formal Models of Action and Cognition

To consider the nature and utility of formal models, we start from basic principles and address the usefulness of a formal approach at various levels. In brief, we first appeal to systems biology to identify the sorts of variables (states) that one needs to consider when modeling an agent immersed in its proximate environment. Equipped with a partition of states, we then utilize optimality principles to define classes of state ("as if") theories of action and cognition; where each state theory is defined in terms of the quantity that is optimized. Finally, each state theory entails a series of process theories that hypothesize a particular (computational or physiological) process that realizes the optimization. Having defined a set of process models, it is then possible to test them in relation to the empirical behaviors that each predicts.

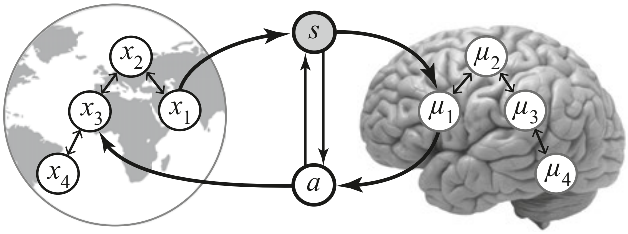
An example of an optimality principle would be Bayes optimality (i.e., ideal Bayesian observer assumptions), where the state theory could correspond to the Bayesian brain hypothesis, in which the brain behaves as if it is trying to maximize Bayesian model evidence. The corresponding process models could then include predictive coding or (stochastic) population coding that make very different predictions about the neuronal responses that would be elicited by a stimulus. We use this example (among others) to see how the models could be augmented to accommodate an action-oriented paradigm.

State Spaces and Systems Biology

Figure 10.2 illustrates the partition of states necessarily implied by a system that is acting on its environment. This partition considers the distinction between external states of the world that are hidden from the internal states of an agent—in the sense that external states are hidden behind sensory states. Internal states could correspond to neuronal activity, connection strengths, or any other neuronal states characterizing the brain at one point in time. Crucially, sensory states are caused by external states that subsequently change internal states. Conversely, internal states cause changes in agential states (e.g., actuators or muscles), which then cause changes in external states. Sensory and

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

$$\partial x / \partial t = f_x(x, a)$$
$$\partial s / \partial t = f_s(x, a)$$



a = agential states
s = sensorial states
x = external states
μ = internal states

$$\partial a / \partial t = f_a(s, \mu) = \partial F(s, \mu) / \partial a$$
$$\partial \mu / \partial t = f_\mu(s, \mu)$$

Figure 10.2 Partition of states necessarily implied by a system that is acting within its environment.

agential states, therefore, couple the world to the brain in a circular fashion, inducing a cycle of action and perception (Fuster 1990). Mathematically, these states insulate the internal states from the external states (and are known technically as a Markov blanket).

The connections or edges in Figure 10.2 denote causal or statistical dependencies and render the graph a Bayesian network or graphical model of the situated agent. These dependencies are most generally described in terms of (stochastic) differential equations. These equations of motion describe the evolution of hidden states, the way that sensory states respond to hidden states and the dynamics of internal states and agential states. The upper two sets of equations constitute a description of the world and how it causes sensory impressions. The lower two sets of equations (for internal and agential states) can now be regarded as a formal model of perception and action.

Optimality Principles

Optimality principles are so ubiquitous in the physical sciences that it is difficult to think of an example in physics that does not rely on an optimality principle. Important examples include Hamilton's principle of least action, which underlies all classical motion and thermodynamic laws, which underlie the behavior of systems at thermodynamic equilibrium in statistical physics. In our case, we can define a state theory of action and perception by casting the equation of motion for internal and agential states as a gradient descent (or ascent) on some function of internal and sensory states. This presupposes that the dynamics of the system or situated agent—say, the sensorimotor loop and/or the neuronal system—can be derived from a variational principle; that is, it

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

assumes we can identify some quantity that the system is trying to optimize. The existence of this quantity (known as a Lyapunov function) is generally guaranteed for the sorts of systems in which we are interested. Identifying the quantity (or quantities) that the system is trying to optimize is crucial because the implicit dynamics (action and perception) will lead to fundamentally different sorts of behavior.

Which Optimality Principle?

There are a range of optimality principles or objective (Lyapunov) functions one might consider. These range from information theoretic quantities based upon the principle of maximum information transfer, or minimum redundancy, through to functions that explicitly accommodate goals, such as utility functions. Some of the more prevalent information theoretic functions are reviewed by Jost (this volume). Most suppose that internal states should possess the greatest mutual information with the hidden or sensory states. In other words, one should be able to predict the external states, given the internal states. An important aspect of these optimality functions is the constraints or priors under which mutual information is maximized. In Pezzulo et al. (this volume), we see a ubiquitous constraint; namely, sensations are caused by a small number of external states at any one time. This *sparsity assumption* can be used to optimize the form of interactions among internal states using a variety of schemes that lead to receptive field properties and architectures that are remarkably reminiscent of functional anatomy.

In optimal control theory and reinforcement learning, the objective function is generally cast as a utility or reward function, also referred to as negative cost. This optimality function speaks more to action than perception, but perception is usually deployed in the context of some state estimation that implicitly appeals to information theory or the Bayesian brain.

The Bayesian brain hypothesis assumes that the optimality function is either Bayesian model evidence or approximations such as variational free energy. Variational free energy is an approximation to the evidence for a model implicit in the internal states that is evident in sensory states at any given time. This means that if internal states minimize variational free energy, they are implicitly maximizing an approximation to Bayesian model evidence and will look as if they are performing (approximate) Bayesian inference, so that internal states come to represent external states.

Many different optimality principles have been proposed to understand human and animal cognition, and the question of which optimality principle to adopt may appear unresolved. Here, we address some of the key issues in this area. For instance, in some schemes, surprise (i.e., prediction error) is minimized, as in the Bayesian brain hypothesis, whereas other proposals emphasize the resolution of uncertainty by maximizing Bayesian surprise or information gain. In robotics or machine learning settings, for example, exploratory

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

actions can be cast as satisfying curiosity (Schmidhuber 1991b). More generally, optimization principles could emphasize the gain of relevant information or the accuracy and scope of predictions. Either could be subordinate to the other. An agent could try to gather information to improve its predictions, or it could build predictions to acquire new information. Jost (2004) argues that a system could satisfy these two goals by addressing them on different timescales. Evolutionary thinking may help us understand the hierarchical relationship between different goals. Short-term predictions provide a reference or set point for homeostasis (or allostasis), whereas relevant information could provide opportunities for reproductive success in the long term, which is not a homeostatic affair. According to evolutionary biology, fitness (as expressed by the actual or expected number of descendants) is the most basic principle. In this setting, survival of an individual (a homeostatic affair) is necessary for, but subordinate to, reproduction. Therefore, from this perspective, the gain of relevant information should be the overarching principle, and homeostasis should be subordinate.

The optimality principles considered above all have particular biases toward perception or action. For example, the Infomax principle and the Bayesian brain hypothesis do not accommodate action, whereas optimal control theory and reinforcement learning do not easily accommodate perceptual inference. Is there any way to integrate these optimality principles into a common framework or state theory? One approach is to recast the optimization of action in relation to reward or utility functions as an inference problem. This is known as *planning as inference*. The advantage of this is that one can gracefully subsume action or policy selection (i.e., planning) and state estimation (i.e., perception) within the same optimality principle; namely, the Bayesian brain. Furthermore, it is relatively easy to show that optimizing internal states with respect to the evidence for a *generative model* implicitly maximizes the mutual information between internal and external states, subject to (prior) constraints inherent in the generative model. This brings us to *active inference* described by Friston (this volume), which is closely related to the notion of *empowerment* (an information theoretic principle that explicitly conditions mutual information on action).

Active Inference

Active inference can be regarded as the action-oriented (enactivist) version of the Bayesian brain hypothesis; it requires *both action and perception* to optimize Bayesian model evidence. Bayesian model evidence (the log probability of some sensory states under a generative model) is also known in information theory as (negative) surprise. This means that approximate Bayesian inference is another way of saying agents act to avoid surprising sensations (i.e., homoeostasis). Clearly this provides an impoverished account of goal-directed and mindful behavior. However, surprise rests on the violation of predictions,

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

where predictions imply prior beliefs about what will happen. Formally, the generative model is specified in terms of likelihood and prior beliefs. Put simply, this means that there are as many classes of optimality as there are prior beliefs an agent might entertain about hidden states and their sensory consequences (e.g., garnering information that may enhance reproductive success). In other words, we can now express any optimality function from the previous section as a prior belief. The only thing that has changed is that the optimality principle is described in terms of a generative model, allowing the same objective function (Bayesian model evidence or its free energy approximation) to be optimized in all cases. Casting optimal control or Infomax principles in terms of active inference does not fundamentally change the principle; it just provides a common framework in which to model the optimization dynamics that underlie action and perception. In short, instead of asserting that agents maximize utility, we can say that agents believe they will maximize utility and then realize those beliefs through action. This emphasizes the fact that utility functions and prior constraints—necessary to define information theoretic imperatives—can be formulated as part of the generative model embodied in the agent's functional anatomy.

One of the most promising and generic priors arises in optimal control theory and decision theory. Known as KL or risk-sensitive control, it simply states that "I act (believe I will act) to minimize the probabilistic difference between preferred outcomes and those I predict given current evidence about the state of the world." The nice thing about this prior is that it gracefully accommodates reward (or utility) and epistemic value (or information gain). In other words, optimizing internal and agential states under this prior leads naturally to a Bayes optimal mixture of explorative and exploitative behavior.

Implications of the Modeling Approaches for a Paradigm Shift

Building on these examples of both formal and informal modeling approaches, we address the implications of these accounts for the role of action in cognition, both conceptually and practically. Areas are highlighted where the modeling approaches provide an account or insight that may differ from other accounts of an action-oriented cognition.

Implication 1: From Open-Loop to Closed-Loop Theories

The first implication that is clear from the system depicted in Figure 10.2 is that the formal account describes a closed-loop scheme. In other words, the external states change the sensorial states, which change the internal states, which change the agential states, which change the external states, etc. In the closed-loop account, perceiving the world produces states that produce action which in turn change the world, causing the loop to iterate. This closed-loop

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

account is distinct from an open-loop account, where the external states change the sensorial states which change the internal states. Note that an open-loop account can be accommodated by the scheme in Figure 10.2 simply by removing the agential state. This has important implications, which we return to later. In the open-loop account, the primary interest is in how the individual responds to the world. The effects on action are largely irrelevant as are effects of action on the world. Action can occur but only as a response that serves to provide data on underlying cognitive mechanisms. The consideration of whether the system is closed or open loop has implications for how cognitive scientists design and interpret data from experiments. In practice, most research in cognitive science and cognitive neuroscience adopts the open-loop approach, in the sense that the paradigms are not constructed to assess closed-loop performance and effects. If cognitive mechanisms have evolved and developed to support closed-loop performance, as suggested, then theories assuming only open-loop processes may be misguided. One potential paradigm shift would be to study cognitive mechanisms in closed-loop paradigms. Although this will present major methodological challenges (e.g., considerable increases in complexity for design, control, data, analysis, etc.), we might find that our understandings of cognitive mechanisms change considerably. One cognitive domain in which such a change from an open-loop to a closed-loop perspective may be especially important is language, which we use to illustrate some key points.

Language as Action

In linguistics and psycholinguistics, language is often considered to be a system of lexical entries and rules that determine the hierarchical structuring of morphological, lexical, and phrasal units (as is evident in speech or text), distinct from its actual implementation in action and perception. Moreover, language comprehension has traditionally been studied separately from production (and the majority of studies concern comprehension with only a smaller number of studies concerning production). However, it is crucial to see that similarly complex combinatorial schemes may be at work in the linguistic and general action domains (Jackendoff 2011; Pulvermüller and Fadiga 2010) and that, when language is used and learned, it is produced and understood in face-to-face communicative contexts, where production and comprehension are intertwined (e.g., gestures, facial, and bodily movements) between conversational partners to fulfill specific communicative goals. These actions complement speech in real-world contexts and are part and parcel with the linguistic signal. The difficulty in separating the linguistic from communicative information becomes especially clear when we consider languages that can only be transmitted in a face-to-face situation, such as sign languages, but it is just as relevant for spoken languages (Vigliocco, Perniss et al. 2014).

Finally, language can be viewed as a tool for communicative action and interaction (e.g., Austin 1975; Searle 1969), and it is indeed the communicative and

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

social context that determines whether an utterance such as "water" functions as a naming action or as a polite request bringing about a desired response. Recent work demonstrates that these different social-communicative functions of language have different brain-mechanistic bases (see Pulvermüller, this volume).

From the perspective of an action-oriented view of cognition, we see immediately that a traditional linguistic definition of language is too narrow. Viewing language as action has a number of important consequences and benefits. First, taking a closed-loop approach to language processing, we can consider language production and comprehension as forms of action and action perception. For example, Pickering and Garrod (2013a) specifically argue that in language processing, speakers construct forward models of their actions before they execute those actions, and perceivers of others' actions (listeners) covertly imitate those actions and construct forward models of those actions. Further, Pickering and Garrod have shown how such a closed-loop approach to language production and comprehension can account for a number of psycholinguistic phenomena, in terms of how speakers/listeners interweave production and comprehension processes as well as how production-based predictions are used to monitor the upcoming utterances in dialogue.

Second, and more broadly, an action-oriented closed-loop view of communication opens up new directions in research that use real-world stimuli where the linguistic content expressed in speech but also co-occurring hands, facial, and body actions are part of the communicative actions. Regarding the role of these additional actions such as co-speech gestures, there is evidence that they are integrated during online spoken comprehension (Kelly et al. 2010; see review by Ozyurek 2014). It has further been shown that they provide a critical cue in vocabulary acquisition such that pointing to objects by infants is a precursor to learning objects' names (Ozçalıskan and Goldin-Meadow 2009).

These different approaches all appeal to the notion that the object of investigation (language) cannot be defined in the traditional reductionist way (rule governed concatenation of symbols). Instead, they should be seen in the context of (and serving the function of) communication. Thus these approaches call for new methods to study language in real-world contexts.

Implication 2: Mental Representations as Inferences about External States

From the biological system account described above and shown in Figure 10.2, one can see immediately that internal states must represent external states. This is because prior preferences about outcomes can only be caused by agential states. However, agential states are only functional in sensory and internal states. This means that internal states must stand in for or represent (in some sufficiency sense) external states. This simple observation dismisses radical accounts of enactivism that preclude (implicit) representations and leads us to a formal account of enactivism: namely, agential stages enact the predictions

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

represented by internal states, where internal states correspond to the production of (conscious or unconscious) inferences about external states based on sensory evidence. A convincing demonstration of the role of generative modeling in perception can be found in Nair and Geoffrey (2006). In brief, they show that when building a model capable of classifying handwritten digits, the inclusion of a generative model of how handwritten digits are created greatly improves classification performance (i.e., perceptual inference). This example speaks directly to the embodied nature of generative models the brain might employ to make sense of the sensations generated by (oneself and) others.

Implication 3: The Key Role of Agential States

From the system biology approach, one can see easily the distinction between enactivist and cognitivist formulations by considering the graphical formulation with and without agential states. If one simply removes the agential states (or action) from the system, one can see that the system is still capable of producing lots of interesting inference and (deep) learning. This would be consistent with the vast literature on perception and cognition that does not rely on active sampling. However, as soon as we place agential states into the mix, we now have the interesting issue of how a Bayesian brain would cope when it can choose the sensory evidence to sample. A key aspect of this is that cognitive attributes, such as the value of information, curiosity, and intrinsic reward, only have meaning in the enactivism paradigm. For example, to address the exploitation-exploration dilemma, one has to account for action. In a similar vein, visual search paradigms would not have any meaning from a purely perceptual or cognitivist perspective. Although it is clear that some inference and cognition can be performed by the system without action, speaking against a pure enactivist account, it is also evident that action has the potential to alter perception and cognition radically. This is perhaps most evident in the ability of the agent to conduct an active search to explore the environment during learning, maximizing the information gain from the senses through acting and moving in the environment. In other words, with action, the agent is able to explore the environment, altering the information about the environment that the internal states can access through the sensorial states. Indeed, there is a large literature from robotics that demonstrates this to be the case (Tsotsos 1992; Shubina and Tsotsos 2010).

There seems to be little disagreement that action is important for endowing artificial agents with cognitive capabilities. Several ideas about how to gain additional information from movement have been explored during the last decades. One example is *active vision*, in which the combination of sensor readings from different viewing angles allows higher recognition accuracy than using each of the single readings (Tsotsos 1992; Shubina and Tsotsos 2010). In robotics, the improvement in performance with active visual search is an existence proof that the action-oriented approach is feasible. More importantly,

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

the behavior of these robots shows what can be accomplished within this paradigm. An earlier and equally impressive example is that of the object recognition strategy implemented by Wilkes and Tsotsos (1992). Here, origami objects piled in a jumble can be individually recognized by a camera mounted on a robot arm that can purposefully move about the pile, selecting viewpoints and object characteristics that are used to isolate and identify them. Related approaches in robotics and computer science could generate new predictions with respect to the paradigm shift toward a more action-oriented view of cognition. For example, it has been shown that the ability to associate behavior with a stimulus is intractable in the general sense without attention (Tsotsos 1995, 2011). This suggests that future theories of attention must be broad enough to handle the requirements of an action-oriented paradigm shift.

Demonstrations of the importance of action for learning and cognition are not limited to robotics and computer vision; they have been also demonstrated in humans and animal models. In humans, recent work on vocabulary acquisition has shown that the learning of labels improves more when the infant actively explores the object being named by a caregiver, than when the child simply looks at the object without actively manually exploring it (Yu and Smith 2013, see also Dominey et al., this volume). In animals, the well-known experiment on vision performed by Held and Hein (1963) dramatically demonstrates the importance of active vision. In this study, a pair of kittens was harnessed to a carousel: one was harnessed but stood on the ground and was able to move around by itself, whereas the other was placed in the gondola and was only able to move passively. The point of this experiment was that both kittens learned to see the world, receiving the same visual stimulation. The difference was that the one could move actively, while the other was moved passively. According to Held and Hein, only the self-moving kitten developed normal visual perception. The other, which was deprived of self-actuated movement, could not develop depth perception. In short, self-movement was necessary to the development of normal visual perception with depth. Our movement in the world, the movement from here to there or there to here, gives the dimension of depth to mere visual sensations. The conclusion is that movement is the key to understanding vision.

Future Opportunities

Real-World Experimentation in Humans

New theoretical frameworks require new experimental paradigms and novel analytical methods. Our predominant methods for studying the brain (e.g., the subtractive approach in fMRI) associate particular areas of the brain with particular functions, but are less informative with respect to how regions form networks and how various networks interact. Methodological approaches need

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

to be formulated so that we can study the activation of simultaneously active neural circuits in the brain in response to naturalistic stimuli. Other necessary advances include:

- Software for making and annotating naturalistic stimuli
- Virtual reality to allow more naturalistic interaction while maintaining experimental control
- Use of mobile measures (eye-tracking, NIRS, EEG)
- Analytic tools for studying interacting brains with fMRI and MEG, and data-constrained modeling based on this data

Experimentation in Robotics

One subtle implication of the formulation offered above is that maximizing expected utility (through pragmatic actions) or epistemic values (through epistemic actions) can be cast as a pure inference problem (using standard Bayesian techniques). This naturally prescribes a space of process theories, each based on different forms of (approximate) Bayesian inference. Practically, this also allows robotic research to avail itself of mature algorithms and schemes that have been considered in great depth over the past decades in statistics and machine learning.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

Bibliography

Note: Numbers in square brackets denote the chapter in which an entry is cited.

Arbib, M. A., A. Billard, M. Iacoboni, and E. Oztop. 2000. Synthetic Brain Imaging: Grasping, Mirror Neurons and Imitation. *Neural Netw.* **13**:975–997. [10]

Austin, J. L. 1975. How to Do Things with Words (2nd edition). Oxford: Oxford Univ. Press. [10]

Fuster, J. M. 1990. Prefrontal Cortex and the Bridging of Temporal Gaps in the Perception-Action Cycle. *Ann. NY Acad. Sci.* **608**:318–329. [10]

Harless, E. 1861. Der Apparat Des Willens. *Zt. für Philosophie und Philosophische Kritik* **38**:50–73. [10]

Held, R., and A. Hein. 1963. Movement-Produced Stimulation in the Development of Visually Guided Behavior. *J. Comp. Physiol. Psychol.* **56**:872–876. [02, 10, 11, 13]

Hommel, B., J. Müsseler, G. Aschersleben, and W. Prinz. 2001. The Theory of Event Coding (TEC): A Framework for Perception and Action Planning. *Behav. Brain Sci.* **24**:849–878. [01, 05, 10, 17]

Jackendoff, R. 2011. What Is the Human Language Faculty? Two Views. *Language* **87**:586–624. [10]

James, W. 1890. The Principles of Psychology. New York: Holt. [07, 10, 14, 17]

Jeannerod, M., M. A. Arbib, G. Rizzolatti, and H. Sakata. 1995. Grasping Objects: The Cortical Mechanisms of Visuomotor Transformation. *Trends Neurosci.* **18**:314–320. [10]

Jost, J. 2004. External and Internal Complexity of Complex Adaptive Systems. *Theory Biosci.* **123**:69–88. [08, 10]

Kelly, S., A. Ozyurek, and E. Maris. 2010. Two Sides of the Same Coin: Speech and Gesture Mutually Interact in Language Comprehension. *Psychol. Sci.* **21**:260–267. [10]

Lotze, R. H. 1852. Medicinische Psychologie Oder Physiologie Der Seele. Leipzig: Weidmannsche Buchhandlung. [07, 10, 17]

Memelink, J., and B. Hommel. 2013. Intentional Weighting: A Basic Principle in Cognitive Control. *Psychol. Res.* **77**:249–259. [10]

Nair, V., and E. H. Geoffrey. 2006. Inferring Motor Programs from Images of Handwritten Digits. In: Proc. of the Neural Information Processing Systems (NIPS 2005), ed. Y. Weiss et al., pp. 515–522. Cambridge, MA: MIT Press. [10]

Ozçalıskan, S., and S. Goldin-Meadow. 2009 When Gesture-Speech Combinations Do and Do Not Index Linguistic Change. *Lang. Cogn. Process.* **24**:190–217. [10]

Ozyurek, A. 2014. Hearing and Seeing Meaning in Speech and Gesture: Insights from Brain and Behaviour. *Phil. Trans. R. Soc. B* **369**:1651. [10]

Pickering, M. J., and S. Garrod. 2013. How Tightly Are Production and Comprehension Interwoven? *Front. Psychol.* **4**:238. [10]

Pulvermüller, F., and L. Fadiga. 2010. Active Perception: Sensorimotor Circuits as a Cortical Basis for Language. *Nat. Rev. Neurosci.* **11**:351–360. [01, 04, 09, 10, 14]

Pulvermüller, F., and M. Garagnani. 2014. From Sensorimotor Learning to Memory Cells in Prefrontal and Temporal Association Cortex: A Neurocomputational Study of Disembodiment *Cortex* **57**:1–21. [09, 10]

Schmidhuber, J. 1991 Curious Model-Building Control Systems. In: IEEE International Joint Conference on Neural Networks, pp. 1458–1463. Singapore: IEEE. [10]

Searle, J. R. 1969. Speech Acts: An Essay in the Philosophy of Language. Cambridge: Cambridge Univ. Press. [09, 10]

Shubina, K., and J. K. Tsotsos. 2010. Visual Search for an Object in a 3D Environment Using a Mobile Robot. *Comput. Vis. Image Underst.* **114**:535–547. [10]

Tsotsos, J. K. 1992. On the Relative Complexity of Active vs. Passive Visual Search. *Int. J. Comput. Vis.* **7**:127–141. [10]

———. 1995. Behaviorist Intelligence and the Scaling Problem. *Artif. Intell.* **75**:135–160. [10]

———. 2011. A Computational Perspective on Visual Attention. Cambridge, MA: MIT Press. [10]

Vigliocco, G., P. Perniss, and D. Vinson. 2014. Language as a Multimodal Phenomenon: Implications for Language Learning, Processing and Evolution. *Phil. Trans. R. Soc. B* **369**:20130292. [10]

Wilkes, D., and J. K. Tsotsos. 1992. Active Object Recognition. In: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 136–141. Urbana: IEEE [10]

Yu, C., and L. B. Smith. 2013. Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS One* **8**:e79659. [10]