14

Consciousness in Action

The Unconscious Parallel Present Optimized by the Conscious Sequential Projected Future

Paul F. M. J. Verschure

Abstract

This chapter outlines the historical cycle of the dominant views in the study of mind, brain, and behavior and the resulting trajectories taken in science. It discusses the grounded enactive predictive experience (GePe) framework, capturing contemporary science and philosophy of consciousness. It advances the hypothesis that consciousness is a necessary ingredient in a behavioral control architecture that has to solve action in a multi-agent world (the H5W problem). Using the distributed adaptive control theory, it shows how apparent heterogeneous approaches can be synthesized to gain greater understanding of a broad range of properties of mind and brain. As the "pragmatic turn" in cognitive science continues to be analyzed, it advocates a reorientation to the study of mind by returning to fundamental issues involved in consciousness.

Introduction

Understanding the nature of consciousness is one of the grand scientific challenges still confronting science today. Fundamentally, the problem involves how to account for phenomenal first-person experience in a third-person verifiable form. Also referred to as the hard problem or the explanatory gap (Levine 1983; Chalmers 1995), it is rooted in the rejection of structuralism by behavioralists about 100 years ago. Interestingly, the scientific study of consciousness has led to two extreme views: (a) as an epiphenomenon (Dennett 1992) or (b) as a fundamental property of matter on a par with mass, charge or space-time (Chalmers 2010; Koch 2012; Tononi 2012). The former perspective deems the phenomenon irrelevant while the latter resorts to panpsychism. This sets up

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.

somewhat of a paradox as the ontology of this phenomenon is placed beyond the scientific method: something to assume rather than to explain. Obviously, a number of alternative proposals fall in between these extremes and aim at finding a link between consciousness and its neuronal substrate (Crick and Koch 1990; for a review, see Dehaene and Changeux 2011). The paradox in the scientific study of consciousness signals a deep conceptual crisis at the heart of phenomenology and psychology: Have we reached the end of the science of mind because we are unable to get past an unsolvable riddle (Horgan 1997)? Or, as summarized by a recent popular newspaper article: “Why can’t the world’s greatest minds solve the mystery of consciousness?” (Burkeman 2015).

The distributed adaptive control (DAC) theory has been widely tested for over twenty years in the domain of both H4W and H5W. DAC has explained a broad range of properties of mind and brain, made predictions that have been corroborated and further elaborated, and allowed for the control of real-world systems ranging from interactive installations and robots to virtual reality-based neurorehabilitation interventions (Verschure 2012b; Verschure et al. 2014). In doing so it has taken an inclusive approach incorporating the core positive values expressed in the mind-brain cycle (Figure 14.1). In short, it links to behaviorism in its focus on embodied action, the core behavioral paradigms of classical and operant conditioning, and the insistence on empirical grounding of knowledge while allowing explicitly defined intervening variables to enter the explanatory framework and avoiding scientism. It incorporates key values

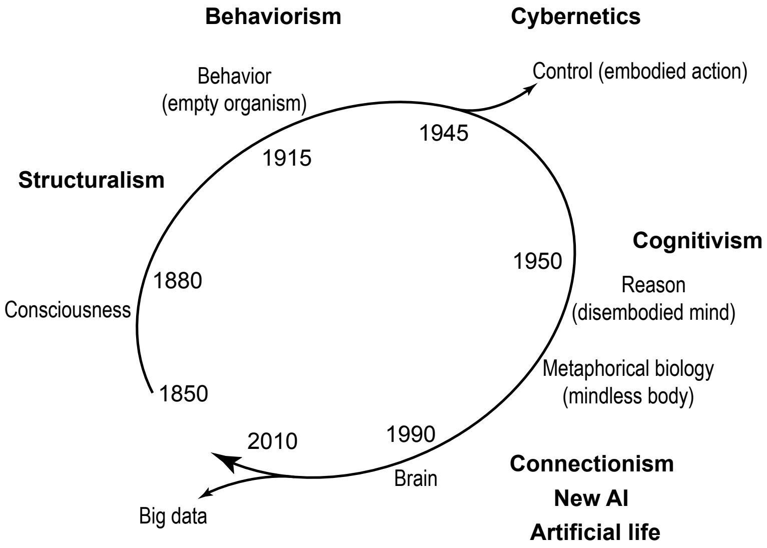

Figure 14.1 The mind-brain-behavior loop starting with experience and performance in the mid-nineteenth century and currently focusing on the brain and its derived data. See text for explanation.

from cybernetics by looking at the mind-brain as an embodied control system that can be studied through synthesis and explains key aspects of reasoning and problem solving as pursued by artificial intelligence while solving the symbol-grounding problem. In addition, it is consistent with the objectives of more recent approaches to anchor the science of mind in that of the brain, staying away from a metaphor. Lastly, DAC does not follow the mirage of big data but rather sees the challenge to get back to the fundamental question of consciousness.

I begin with a small historical detour to demonstrate that the situation in which we find ourselves is no accident but rather a logical consequence of the trajectories that the study of mind has followed over the last 150 years of scientific enquiry. I will then use this to propose a reorientation of the study of mind, in particular consciousness, toward a synthetic and action-oriented paradigm. I argue that consciousness is a transient memory system that functions specifically to mediate between the self and the world, providing valuation of parallel control systems and being causal with respect to future action.

The Mind-Brain-Behavior Loop

The scientific[1] study of mind and brain has followed a very specific development of concepts and methods. The main phases of this process have impacted a range of disciplines including psychology, neuroscience, computer science, linguistics, and philosophy. Our current situation in terms of the study of mind and the renewed focus on consciousness and action can best be understood if it is grounded historically. This will allow us to identify with greater clarity the novel contributions to the discussion of the pragmatic turn in the study of mind and brain and distinguish it from repetitions, redundancies, and noise (for reviews, see Koch and Leary 1985; Gardner 1987; Pfeifer and Scheier 1999).

Structuralism and the Primacy of Subjective Experience

The scientific study of mind began in the second half of the nineteenth century with the continental psychology of Fechner, Helmholtz, Donders, and Wundt. Wundt's structuralism, seen as the first school in psychology, was characterized by the systematic scientific study of phenomenology and human performance. Structuralism tried to combine rigorous empirical methods with the study of instantaneous experience through introspection. Wundt gave precedence to free will over reason, or voluntarism, and advocated a combined passive associative and active interpretative or "apperceptive" process in the construction of experience. This approach, among other things laid the foundation of

[1] I adhere to the restricted definition of science as the attempt to ground knowledge in direct observation.

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.

modern psychophysics and sequential-processing models of the mind (Miller in Koch and Leary 1985). Hence, the initial scientific study of psychology took conscious experience as its explanandum.

Behaviorism and the Empty Organism

Behaviorism emerged in the first half of the twentieth century, driven by the confluence of (a) the comparative study of behavior set forth by the Darwinian revolution, (b) the pragmatism of Peirce, James, and Dewey, which grounds knowledge in its practical outcomes, and (c) what we can call "physics envy" or scientism. Behaviorism was a direct reaction to structuralism, negating its core dogmas and advancing three fundamental ideas (Kendler in Koch and Leary 1985): translate the methods of the natural sciences to the study of behavior; rely on the study of psychology on behavior as the dependent variable; and reject mental states on methodological grounds. In adopting Bridgman's physics-inspired philosophy of operationalism, the behaviorism of Watson and Skinner advanced the notion of an empty organism: a passive object fully controlled by the external forces of its world and explained in terms of the reflex atom.

Although Bridgman's model turned out to be a caricature of late nineteenth century physics, the drive to force the conceptualization of the phenomena under study into the methods employed became a prime example of *scientism*, or the excessive need for quantification even in domains that do not satisfy its specific conditions (Hayek 1943). This empiricist perspective on knowledge, formalized by the logical positivism of the time (which sought a link between logic and observation), also served the *unity of science* agenda; the psychology of adaptive behavior could be reduced to the biology of the brain, which would in turn give way to explanations at lower levels of description (i.e., chemistry and physics).

After about half a century of trying, behaviorism failed to deliver on its promise of identifying universal principles of adaptive behavior. Behaviorism also did not manage to scale up to more advanced forms of behavior beyond salivating, twitching, freezing, pushing levers, or pecking. Most importantly, organisms were not enslaved by the reinforcement received from the environment, as the empty organism dogma prescribed. Instead, autonomously structured learning and behavior was dramatically demonstrated in the experiments of Tolman, for example, leading to the notion of the cognitive map (Tolman 1948) and the Rescorla and Wagner laws of classical conditioning: animals only learn when events violate their expectations (Rescorla and Wagner 1972). Behavior could just not be explained by ignoring the agency and self-structuring of adaptive behavior by the organism itself.

Behaviorism negated structuralism and its explanandum, consciousness, thus giving precedence to the empirical methods deployed in the study of mind, which gravitated to the ultimate dependent variable: behavior. By placing method before concept, however, behaviorism went down the rabbit hole of scientism.

The Disembodied Mind of Cognitivism

The crisis that ensued through the collapse of behaviorism was resolved by switching to the computer metaphor of mind that gave rise to artificial intelligence (AI) and the cognitive science of the second half of the twentieth century, spearheaded by McCarthy, Newell, Simon, and Minsky, and the linguistics work of Chomsky (for a review, see Gardner 1987). AI won out over cybernetics, the latter placed emphasis on control and real-world action, whereas the former advanced a study of the logical operations performed by a disembodied rational mind. The neo-functionalism that followed (i.e., mind as explained in terms of rules and representations rather than its substrate or operant expression) argued for a unique level of explanation akin to that of a Turing machine (Putnam 1960). Hence, the move toward the computer metaphor negated the dogmas of behaviorism and its link to the rigorous empirical investigation at the level of brain and behavior—logically advancing the notion of a special level of explanation, yet isolated from implementation by virtue of multi-instantiation. As a result, the notion of an algorithmic computational reasoning system replaced that of mind as part of an embodied acting system, as demonstrated by the General Problem Solver of Newell and Simon (1963) and the SOAR cognitive architecture (Newell 1990).

Unfortunately, cognitivism and AI, in turn, stumbled over its claims of being able to both explain and synthesize intelligence. This intellectual failure combined with a lack of impact in the real world led to the so-called AI winter, during which funding and interest evaporated. As a research program, AI got bogged down in symbolic grounding (Harnad 1990) and the frame problem (McCarthy and Hayes 1969), both of which relied critically on the prior specification of the rules and representations that purportedly should have explained "intelligence" (Verschure 1998). In other words, AI's successes were largely due to designers defining appropriate priors into their systems as opposed to these artificial intelligences autonomously acquiring them, also referred to as the *problem of priors* (Verschure 1998).

Hence, the cognitive revolution, and its AI spearhead, sacrificed the empirical methods of behaviorism to develop a *new* science of reason decoupled from the natural science approaches, capitalizing on synthetic methods afforded by the emergence of the computer. This disembodied mind and its associated theoretical framework, however, was more a reflection of the designers' fancy coupled with technical capabilities than of any kind of natural system.

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.

A Metaphorical Biology of Mind and Brain

Although a historical account of the recent past concerning the study of mind and brain requires a perspective that only time will provide, clear trends can already be distinguished (for an optimistic initial rendering, see Pfeifer and Scheier 1999). The disembodied mind of AI was followed in the late 1980s by a period of research in which biological metaphors guided the study of mind and brain, based partially on the construction of artificial systems. Examples include behavior-based AI, new AI, artificial life, genetic algorithms, neural networks and connectionism (Pfeifer and Scheier 1999) combined with a philosophy of "eliminative materialism," where the whole of human experience would be described in "brain speak" (Churchland 1986) that harked back to the philosophical behaviorism of Ryle (1949). "New" AI directly negated its predecessor (symbolic AI) by proposing a nonrepresentational, behavior-based, and embodied explanation of mind, whereas connectionism sought out the "subsymbols" that would link mental states to the neuronal substrate. Neither approach had a lasting conceptual impact on the study of mind and brain beyond signaling a shift toward new methods, such as those used in computational neuroscience, embodied cognition, or biomimetic robotics. However, this methodological advance, which opened up a universe of *in silico* experimentation, was realized at the expense of an organizing theoretical framework and a clear coupling back to empirical science.

Big Data and Avoiding Conceptual Defeat

The emergence of big data at the beginning of the twenty-first century exemplifies the regression of the study of mind and brain: the idea of advancing a theory was explicitly abandoned in favor of collecting large amounts of data and is well exemplified by the large-scale brain-oriented projects currently being pursued in the United States and Europe (Verschure 2016). From data-free theory, we fall into the antithesis of theory-free data. Placed in historical context, big data appears to be the logical result of the preceding period: failing to crack the conceptual problem of mind and brain over the previous two centuries, we resort to the last viable vestige of the scientific method—the collection of data. Big data emerges at a critical point in the study of mind and brain and can be seen as signaling a scientific crisis (Horgan 1997; Kuhn 1962/1970).

This crisis has been further amplified through the contemporary trend of studying consciousness outside the scope of science, due to the exceptional status attributed to consciousness: it was essentially removed from the scientific agenda by Dennett (1992), who holds that it is epiphenomenal (for a discussion, see Robinson 2010); by Rosenthal (2008), who advocates that it has no function; and by Chalmers, Tononi and Koch, who want us to believe that it is neo-panpsychistic, part of the fundamental fabric of nature (Chalmers 2010; Tononi and Koch 2014). It is indeed ironic that after distinguishing the easy from the hard problem, the latter was solved by assumption and declared not to be a problem at all! This can be viewed as a form of explanatory nihilism (Price 2002):

> What shall we do? Many would find relief at this point in celebrating the mystery of the Unknowable and the "awe," which we should feel at having such a principle to take final charge of our perplexities. Others would rejoice that the finite and separatist view of things with which we started had at last developed its contradictions, and was about to lead us dialectically upwards to some "higher synthesis" in which inconsistencies cease from troubling and logic is at rest. It may be a constitutional infirmity, but I can take no comfort in such devices for making a luxury of intellectual defeat. They are but spiritual chloroform. Better live on the ragged edge, better gnaw the file forever! (James 1890/1950).

This short analysis shows that in five easy steps the study of mind and brain has sacrificed its explanandum, its methods and theories and an intellectual limbo has resulted wherein the questions have shifted from the explanation of psychological constructs to the description of neuronal correlates and its underlying data. These steps are easy because each paradigm followed as a negation of the premises underlying the preceding one, terminating in a science that is about data as opposed to ideas. What shall we do?

Following James's recommendation, I argue that an alternative is to embark on a new study of mind and brain; let us use the unpopular notion of psychology that addresses the fundamental question of consciousness and combines the strength of preceding approaches, as opposed to their overstated promises, while answering their weaknesses. This explains the idea of a mind-brain cycle depicted in Figure 14.1: back to experience as our explanandum! Essentially we need to focus on explaining consciousness, linking it to overt behavior and reasoning based on rigorous empirical, formal, and synthetic methods, and grounding this explanation in the biological principles that govern bodies and brains. Such a program is not necessarily incompatible in its realization with aspects of the approaches listed above. The big difference, however, is that it steps away from the brink of nihilism and declares consciousness, yet again, a phenomenon to be explained, hypothesizing a distinct function and an augmented method to investigate it. This approach is grounded in the *distributed adaptive control* (DAC) theory of mind and brain that has been advanced using embodied biologically grounded models linking the neuronal substrate to action (Verschure et al. 2003; Verschure 2012b). The DAC program realizes what this Forum seeks: linking mind to action.

Instead of assuming that consciousness is a fundamental property of the physical world, an alternative and more straightforward hypothesis (which has not yet been exhausted) is that consciousness is a unique feature of a subset of living systems: it is the product of biology rather than physics as advocated, for instance, by Searle (1998). This means that we have to place its study in the context of evolution to follow Dobzhansky (1973) and consider its function in terms of function and fitness. From this perspective, two features stand out and seem paradoxical: consciousness is defined in terms of one coherent unitary scene (James 1890; Bayne 2010), yet experimental evidence shows that this conscious scene is experienced with a significant delay, relative to the real-time action of the agent (Libet 1985; Haggard et al. 2002; Soon et al. 2008), that is not necessarily the cause of action and thought (Wegner 2003; Custers and Aarts 2010). The resulting paradox is that in optimizing fitness, evolution appears to have rendered solutions to the challenge of survival which include putatively epiphenomenal processes like consciousness.

I propose a solution to this paradox and advance the hypothesis that consciousness is a necessary ingredient of a behavioral control architecture that has to solve action in a multi-agent world, or the, so-called, H5W problem. Before turning to it, I outline the most dominant views on consciousness and show how these can be integrated into one coherent framework, which serves as a context from which we will launch the DAC theory of consciousness and H5W.

The GePe Framework

It has become standard to acknowledge that there are addressable and non-addressable problems in the study of consciousness, or easy and hard problems. With respect to the "easy" problems, a number of core principles underlying consciousness and qualia have emerged. These can be summarized in the *grounded enactive predictive experience* (GePe) model of consciousness (Verschure 2012b, 2013) that will guide model construction and validation. The GePe model utilizes five principles:

GePe 1: Consciousness Is Grounded in the Experiencing of the Physically and Socially Instantiated Self

Experience requires a self that does the experiencing (Nagel 1974; Metzinger 2003; Edelman 1989; Craig 2009). For instance, Edelman (1989) proposed primary and secondary forms of consciousness that relate to the expanding temporal horizon of the self, from the instantaneous physical experience (primary) to the imagined future and remembered past (secondary). Metzinger (2003) refined this notion further: the self progresses from a globalized identification, with the body or first-person perspective, to a transparent spatiotemporal self-localization in the world or minimally phenomenal self based on a form of representation of the self, to a fully fledged phenomenal first-person perspective or strong first-person perspective. The first-person perspective begins as a point of convergence of sensory (but also proprioceptive and interoceptive) experience; then it coalesces into a strong form where the self is internally represented as reflecting the organization of the body and its sensorimotor coupling to the world (see GePe principle 2); this is followed by the representation of the object- and action-directedness of the self (i.e., intentionality) found in the strong first-person perspective. As interactive and social dynamics are rich sources of sensory experience and feelings, they become part and parcel of the representation of the self, which is not only physically but also socially instantiated (Frith 2008). In a sense, this view on self and consciousness also reflects a trend in cognitive science to ground knowledge and experience in embodiment, situatedness, and interaction dynamics (Verschure et al. 1992; Pfeifer and Bongard 2006; Barsalou 2008). Damasio (2012) has recently advanced a similar proposal, retracting his version of the James-Lange theory of emotional experience to suggest now that consciousness requires representations of self to enter into memory.

GePe 2: Consciousness Is Defined in the Sensorimotor Contingencies of the Agent in the World

In neuroscience, cognitive sciences, and robotics there is a shift from a representation-centered framework toward a paradigm that focuses on the intimate relation between perception, cognition, and action (see above). Although many proponents have supported such an "action-oriented" paradigm over the years, starting with Pavlov (Pavlov 1927; Verschure 1992) and his mentor Sechenov, it has only recently started to regain traction. In this view, cognition is not isolated from action and a database-serving planning in a strict sense-think-act cycle, as already advanced by Donders in the nineteenth century. Rather, cognitive processes are closely intertwined with action and can best be understood as "enactive," as a form of practice itself (Pulvermüller and Fadiga 2010; Verschure et al. 1992). The intrinsic action-relatedness of cognition is the core consideration of the *sensorimotor contingency theory* put forward by O'Regan and Noë (2001) that addresses the fundamental role of action for perception and awareness. Accordingly, the agent's sensorimotor contingencies are law-like relations between movements and sensory inputs which provide the foundations for knowledge and experience. O'Regan (2011) has proposed that these laws of sensorimotor contingencies define the qualia of conscious experience. A challenge for this framework, as for its behaviorist ancestors, is to scale-up to cognition, affect, and rich experiences which might appear non-motor.

GePe 3: Consciousness Is Maintained in the Coherence between Sensorimotor Predictions of the Agent and the Dynamics of the Interaction with the World

The idea that perception is defined by predictive models of environmental causes of sensory input enjoys a rich pedigree that extends back at least as far as Helmholtz and the seminal work of Tolman (1932). Indeed, sensorimotor contingencies not only exist instantaneously, they can also be predicted by virtue of their invariance (Bar 2007). It has been proposed that cognition and consciousness are based on such internal simulations of the possible scenarios of interaction with the world using *forward models* (e.g., Hesslow 2002; Cisek 2007). Indeed, it has been proposed that concepts themselves can be seen as simulations (Barsalou 2008). It is through simulation that an "internal" world can appear in consciousness, freeing the organism from its immediate physical environment (Hesslow 2002; Revonsuo 2006). Merker (2005) has argued that the simulated internal world compensates for the uncertainties generated by the dynamics of sensory states due to self-induced motion and that it can be seen as a self-generated virtual reality (Revonsuo 1995).

That the brain is organized around prediction has reached recent prominence in the Bayesian Brain and "predictive coding" frameworks (Bar 2007; Clark 2013b; Verschure et al. 1992; Rao and Ballard 1999; Friston 2005; Barsalou 2008), and was anticipated by Massaro (1997) in his analysis of speech perception. In these views, core structures of the brain (including the thalamocortical and corticobasal ganglia systems as well as the cerebellum) are engaged in hierarchical Bayesian inference, extracting generative models of both sensory inputs and the consequences of action across multiple timescales and modalities (Hesslow 2002; Lau and Rosenthal 2011). In neurophysiological terms, "top-down" connections are suggested to convey the content of these generative (predictive) models, whereas "bottom-up" signals convey prediction errors (Bar 2007; Mathews and Verschure 2011). The resulting models have received growing support (Friston 2005). However, the exact relation between predictive processing and biological consciousness remains poorly understood, although some correlate of this view has been reported in coma patients (Boly et al. 2011). For example, there is no consensus on which sorts of predictive model give rise to conscious contents, and which do not. Furthermore, it is unclear what the relations are between probabilistic representations postulated by the Bayesian brain and the fact that (apparently) we do not perceive our conscious states as being probabilistic. Merker (2005) argues that information, while in cortex, is generally maintained in the form of probability distributions yet the content of consciousness is linked to the "collapsing" of probability functions into a simpler format (hence the reason why our conscious percepts do not appear to be probabilistic). This format is proposed to be required for subcortical processing (Ward 2011) to provide global best estimates of variables of interest within narrow time windows. However, this hypothesis remains to be tested and compared with other mainstream theories that view cortex as the locus of consciousness.

A variation on the prediction-based theories on consciousness is the attention schema theory proposed by Graziano (2013). In this proposal, underlying consciousness is the process of attention; its role of identifying subsets of sensory information of relevance to an agent is attributed to it by an observer. Consciousness is thus seen as the ascription of such attentional states to others and the self.

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.

GePe 4: Consciousness Combines High Levels of Differentiation with High Levels of Integration

More progress has been made with respect to another structural property, namely *complexity*. Following Edelman and Tononi, it is a deeply significant fact that each and every conscious scene is both *integrated* (or unitary) and massively *differentiated*, such that it provides for a highly informative discrimination among a very large repertoire of possible experiences (Edelman 1989; Tononi and Edelman 1998; Tononi 2008, 2012). Viewing consciousness from the perspective of such integrated information suggests theoretically grounded and empirically applicable quantifications of consciousness, such as information theoretic measures (Tononi and Edelman 1998) or multivariate autoregressive modeling (*causal density*) (Seth 2009). Tononi's *integrated information theory* (IIT) provides a precise definition of information integration given a number of assumptions on how to segment informational spaces. IIT introduces a fundamental quantity, integrated information (Φ), expressed in bits, which measures to what extent a system integrates information as a whole via its causal dynamics, over and above that of its subparts. For systems composed of largely independent modules, Φ is low as it is for nonmodular systems that are connected in a homogeneous or random manner. IIT is high only for systems that are both functionally specialized and integrated. This measure has been used to distinguish between different levels of consciousness, where sleep states show a lower complexity than awake and alert states (Massimini et al. 2009). However, this in itself is also a possible drawback because one can confuse the measure with the ontology of consciousness. For instance, what is the bound on Φ, and what are its discriminative capabilities? The notion that complexity reflects consciousness can lead to a misunderstanding of the ontological significance of these measures such that any complex system (e.g., the Internet) must be conscious by definition, leading to the aforementioned panpsychism (Koch 2012). Within this perspective, consciousness does not have a declared function.

GePe 5: Consciousness Depends on both Highly Parallel, Distributed Implicit Factors and Metastable, Continuous, Unified Explicit Factors

Theories of consciousness are tightly constrained and informed by evidence regarding unconscious processing (Baars 1988). In Baars's "global workspace" architecture, specialized unconscious processors compete for access to a central resource: the conscious global workspace. Accordingly, consciousness is ascribed to content that is received from and broadcast back to a broad network of unconscious modules or processors. In this way consciousness provides a serial and integrated stream of qualia that are produced by many subconscious "processors." The key parameter that defines whether content becomes conscious is the ability to penetrate many of these processors. In this respect the global workspace is an example of *access consciousness* (Block 2007). The integration and serialization provided by the global workspace provides for behavioral flexibility by allowing unconscious processors to generate fast responses in familiar situations, while in novel situations the integrated qualia that are broadcast from the global workspace can facilitate the production of new responses (Baars 1988). The global neuronal workspace hypothesis proposes that the workspace comprises perceptual, motor, attention, memory, and value areas which form a common higher-level unified information space that serves a similar role as the global workspace largely dependent on the specific anatomy of corticocortical projections (Dehaene et al. 1998; Dehaene and Changeux 2011). The main function ascribed to the global neuronal workspace is that of assisting in problem solving and executive control (Dehaene 2014).

Epiphenomenalism and the Case against Free Will

Whereas Descartes places phenomenal subjective states at the center of mental existence, a number of converging lines of evidence show that humans are largely unaware of the causes of their own thoughts and actions (Wegner 2003). This observation, corroborated by a large set of experiments, has fueled the interpretation that consciousness is an epiphenomenon; that is, it is an evolutionary leftover with no operational relevance (Dennett 1992). A large amount of cognitive processes can be performed without reportable awareness of the relevant stimuli or contingencies, and some processes (e.g., overlearned motor responses) are supposedly even more effective when implemented by unconscious systems (Baars 1988; Milner and Goodale 1995). The latter claim of an unconscious thought advantage has been put in doubt in a recent meta-analysis (Nieuwenstein et al. 2015). Less appreciated but equally fundamental is the notion that motor actions and intentions can be unconscious as well as conscious (Dijksterhuis and Bargh 2001; Frith et al. 2000a), that unconscious intentions are known to reliably precede conscious awareness of motor actions (Desmurget and Sirigu 2009; Libet 1985), and that behavioral goals can be set by unconscious factors (Custers and Aarts 2010). Another category of implicit factors in experience and action are emotions. Indeed, emotion and consciousness are tightly coupled, and conscious experiences generally involve affective (emotional) components, both transiently (e.g., delight, surprise) and as a background mood (e.g., sadness, contentment, anxiety) (Tsuchiya and Adolphs 2007). Since James-Lange it has been suggested that emotions arise as perceptions of bodily states (Critchley et al. 2004) and that autonomic signals can reflect implicit reactions to salient stimuli, including prediction errors (Uhlhaas et al. 2009). It has further been argued that the processes underlying volitional behavior (e.g., implicit learning, evaluative conditioning, unconscious thought) are intrinsically goal dependent, requiring forms of attention while operating outside of awareness (Dijksterhuis and Aarts 2010). In all cases, conscious and unconscious processes are closely coupled and interact strongly in

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.

generating the stream of consciousness and adaptive behavior (Baumeister et al. 2011). They can be seen as complementary since unconscious processing can be sensitive to patterns, regularities, and other structures within signals prior to conscious awareness, suggesting that the content of consciousness is biased and based on unconscious factors (Baars 1988; Haggard and Eimer 1999). Such dual-process theories (Evans 2008)—such as fast and slow processes in decision making (Kahneman 2011) and the distinction between reasoning, planning, and monitoring processes (Gazzaniga 2011)—face the fundamental question of how these processes are maintained in isolation and interfaced as well as how the exchange of information between them is regulated. In particular, we need to know whether these multiple processes are coherent or are descriptions of a further heterogeneous set of subsystems, possibly leading to an infinite regress (Evans 2008).

Distributed Adaptive Control: A Theory of the Mind, Brain, Body Nexus

The perspectives on consciousness and its putative functions outlined above can come across as rather heterogeneous. However, when each are viewed as highlighting specific and complementary aspects of consciousness and its function, they can be brought together and reconciled. I have synthesized this from the perspective of the distributed adaptive control, illustrated in Figure 14.2, and begin with a brief explanation of the DAC principles.

A highly abstract representation of the DAC architecture is depicted on the left-hand side of Figure 14.2, wherein the brain is organized as a layered control structure with tight coupling within and between the somatic, reactive, adaptive, and contextual layers. Across these layers a columnar organization exists to process the states of the "world" or exteroception (left column), "self" or interoception (middle column), as well as "action" (right column), which mediates the previous two. The somatic layer equips the body with its sensors, organs, and actuators. The reactive layer is made up of dedicated behavior systems which combine predefined sensorimotor mappings with drive reduction mechanisms predicated on the needs of the body (somatic layer).

Depicted in the right lower panel of Figure 14.2 (allostatic control) we see that each behavior system follows homeostatic principles supporting the self essential functions (SEFs) of the body (somatic layer). To map needs onto behaviors, the essential variables served by the behavior system have a specific distribution in space called an affordance gradient. In this example, we consider the (internally represented) "attractive force" of the home position supporting the *security* SEFs or of open space defining an *exploration* SEFs. The values of the respective SEFs are defined by the difference between the sensed value of the affordance gradient (red) and its desired value given the prevailing needs (blue). The regulator of each behavior system defines the next action so

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.

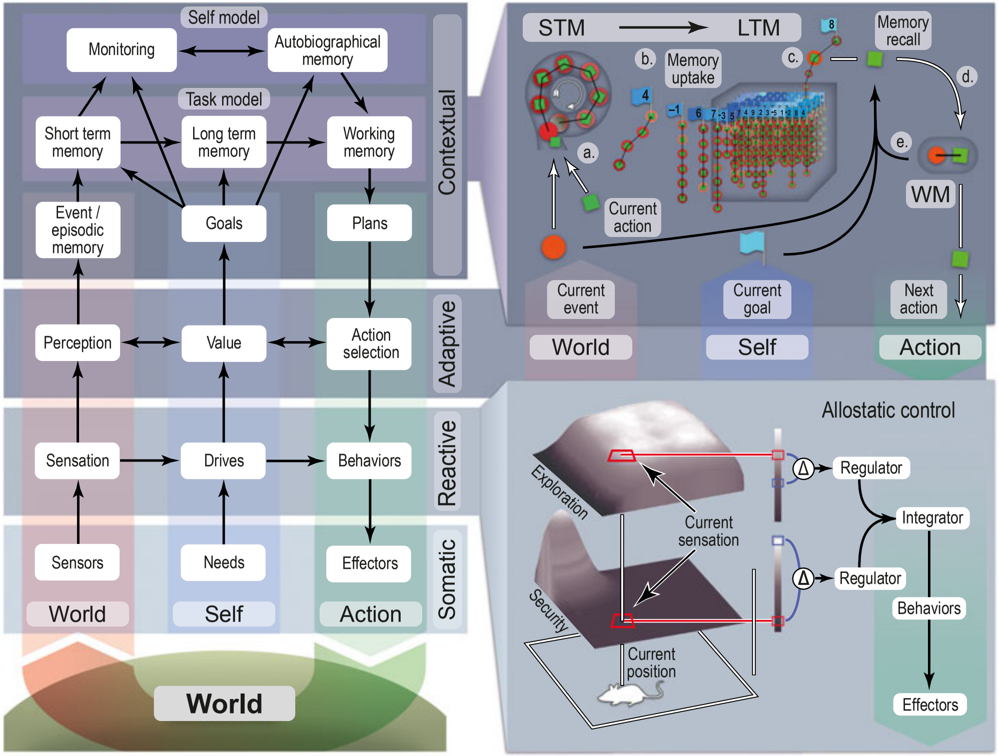

Figure 14.2 The distributed adaptive control (DAC) theory of mind and brain. See text for details. Figure adapted from Verschure et al. (2014); see text for explanation. For an overview of the mapping of DAC principles to the brain, see Verschure (2012b).

as to perform a gradient ascent on SEFs. An integration and action selection process across the different behavior systems forces a strict winner-take-all decision, which defines the specific behavior emitted. The allostatic controller of the reactive layer regulates the internal homeostatic dynamics of the behavior systems to set priorities defined by needs and environmental opportunities through the modulation of the affordance gradients, desired values of SEFs, and/or the integration process. The adaptive layer acquires a state space of the agent-environment interaction and shapes action. The learning dynamics of the adaptive layer is constrained by the SEFs in the reactive layer which define value. Crucially, the adaptive layer contributes to exosensing by allowing the processing of states of distal sensors (e.g., vision and audition). These are not predefined; instead they are tuned in somatic time to properties of the interaction with the environment. In turn, acquired sensor and motor states are associated through the valence states signaled by the reactive layer. The adaptive layer is modeled after the paradigm of classical conditioning (Pavlov 1927), and the acquisition of the sensorimotor state space is based on predictive mechanisms to optimize encoding and counteract biases due to behavioral feedback. The adaptive layer has been mapped to the cerebellum, amygdala, cortex, and hippocampus. The contextual layer is divided between a "world" and a "self" model. It expands the time horizon in which the agent can operate through the use of sequential short- and long-term memory (STM and LTM, respectively) systems. These memory systems operate on the integrated sensorimotor representations generated by the adaptive layer and acquire, retain, and express goal-oriented action regulated by the reactive layer. The contextual layer comprises a number of interlocked processes (right upper panel):

a. When the error between predicted and encountered sensory states falls below an STM acquisition threshold, perceptual predictions (red circle) and motor activity (green rectangle) generated by the adaptive layer are stored in STM as a *segment*. The STM acquisition threshold is defined by the time-averaged reconstruction error of the perceptual learning system of the adaptive layer.
b. If a goal state (blue flag) is reached (e.g., reward or punishment), the content of STM is retained in LTM as a sequence conserving its order, goal state, and valence marker (e.g., aversive or appetitive), and STM is reset so that new sequences can be acquired. Every sequence is thus defined through sensorimotor states and labeled with respect to the specific goal it pertains to and its valence marker.
c. If the outputs generated by the reactive and adaptive layers to action selection are below the threshold, the contextual layer realizes its executive control and perceptual predictions generated by the adaptive layer are matched against those stored in LTM.
d. Action selected in the contextual layer is defined as a weighted sum over LTM segments.

e. The contribution of LTM segments to decision making depends on four factors: perceptual evidence, memory chaining, the distance to the goal state, and valence. The working memory (WM) of the contextual layer is defined by the memory dynamics that represents these factors. Active segments in WM that contributed to the selected action are associated with those which were previously active in establishing rules for future chaining.

The core features of the contextual layer have been mapped to the prefrontal cortex. The self-model component of the contextual layer monitors task performance and develops (re)descriptions of task dynamics anchored in the self. In this way, the system generates meta-representational knowledge to form autobiographical memory.

To create structure in the tangle of neuronal processes and subprocesses that make up the brain and its multilevel organization, we need to define unambiguously what the overall function of this system is. DAC follows Claude Bernard and Ivan Pavlov in defining the brain as a control system that maintains a metastable balance between the internal world of the body and the external world through action. This pertains both to its physical and informational needs. The question thus becomes: What does it take to act?

The DAC theory proposes that to act in the physical world, the brain needs to optimize a specific set of objectives which are captured in answering the questions: *Why* do I need to act? *What* do I need? *Where* and *when* can this be obtained, and *how* do I get it? Embedded within these questions is a complex set of computational challenges that has been termed the *H4W problem* (Verschure 2012b). In short, an agent needs to determine a behavioral procedure to achieve a goal state (the *how* of action). This, in turn, requires defining the motivation for action in terms of needs, drives, and goals (the *why*); the objects and their affordances in the world that pertain to these goals (the *what*); the location of objects in the world, the spatial configuration of the task domain and the location and confirmation of the self (the *where*); and the sequencing and timing of action relative to the dynamics of the world and self (the *when*). DAC theory proposes that goal-oriented action in the physical world emerges from the interplay of these different processes subserving H4W.

Each of the *W*s can be seen as a specific objective that the brain must satisfy. In turn, each can be decomposed into a large set of sub-objectives of varying complexity organized across different levels and scales of organization of the central nervous system. At a first level, the brain must assess the motivational states derived from homeostatic self-essential variables defined at the level of the soma and reactive control. These motivational states, in turn, need to be prioritized so that goals can be set: this is the *why* problem, requiring the modulation of associated behavior systems. Next, a second layer of control is called upon to classify, categorize, and valuate states of the world, to identify the spatial layout of the task, including the agent itself, and the dynamics of the task and its affordances: *what*, *where*, and *when* also engages the learning systems of the adaptive layer. These labeled multimodal states are grouped in sequences around prioritized goals at the level of contextual control; for example, in a rodent navigation set-up, to go toward and push a lever placed at the northeast corner of the environment, given that the cue signal has appeared. At this stage the *how* has been generated and expressed. Using the accumulated spatiotemporal knowledge of the task and the self in which goal pursuit is framed, a procedural motor strategy (*how*) can be composed and its elements selected from the set of available options to achieve a goal state (Verschure et al. 2014).

The H4W framework is an exclusive set of processes that directly maps onto the functions of the different layers of DAC, capturing core brain mechanisms that mediate and control instrumental interaction with the physical world as in the adaptation to an open field (Figure 14.3) or to foraging tasks including neocortex, hippocampus, basal ganglia and the cerebellum (for a review, see Verschure et al. 2014). To solve H4W, we have constructed an architecture that comprises all components of GePe: an embodied self, generating and acquiring sensorimotor contingencies, relying on forward models, displaying the integration of information and maintaining a global workspace in its memory systems (Figure 14.3). This DAC H4W realization has been tested on a range of robots and shows all signatures of GePe; however, it is not conscious, contrary to claims that even simpler models can be called conscious (Tononi and Koch 2014). The reason for this, I propose, is that a fundamental aspect is missing; namely, the ability to simulate *hidden* states of the external world.

H4W solely addresses the interaction of an agent with its physical world. DAC theory proposes that the Cambrian explosion (ca. 550 million years ago) created environments dominated by one more critical factor, which demanded a specific objective function: *Who* is acting? The resulting move from the H4W to the H5W problem leads to a fundamental change in information processing: reciprocity and hidden states.

Reciprocity results from a behavioral dynamic: the agent is now acting on a world that is, in turn, acting upon it. The states of other agents, which are predictive of their actions, are however, hidden. At best they can be inferred from incomplete sensor data, such as location, posture, vocalizations, or social salience (Inderbitzin et al. 2013). As a result, the agent must unequivocally assess, in a deluge of sensor data, those extero- and interoceptive states that are relevant to ongoing and future action. In addition, the agent must deal with the ensuing credit assignment problem to optimize its own actions. In this partially observable intentional world, the solution to survival entails assessing (a) the relevant (hidden) states of the world and its agents, (b) the relevant states of self, and (c) the specific action which gave rise to relevant outcomes. I propose that consciousness is a necessary component of the control system that solves this H5W problem.

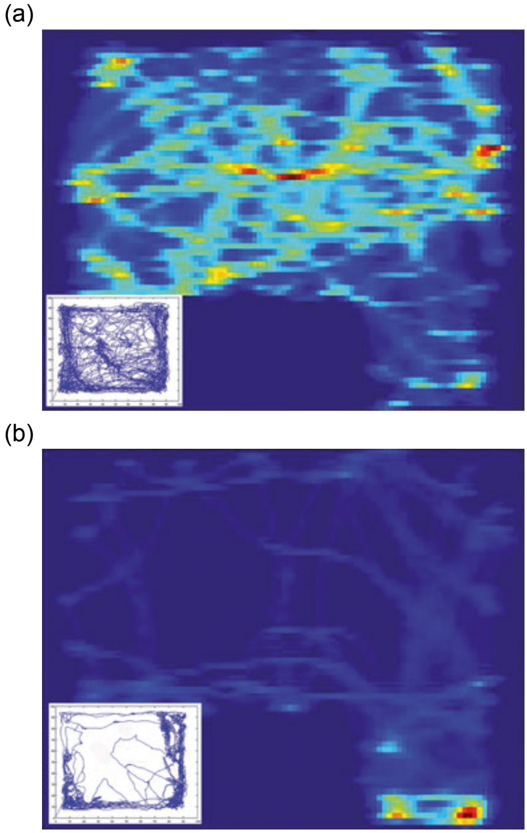

Figure 14.3 H4W solved by DAC in the real world. Density plots are shown of the positions visited by a robot controlled by the DAC architecture while (a) familiar or (b) novel environments are explored. Behavioral trajectories are modulated by the behavioral subsystems of exploration and security respectively. Insets show example trajectories from rats which performed under similar conditions. Adapted from Sánchez-Fibla et al. (2010).

The control system generated by DAC-based consciousness incorporates all elements of the GePe principles and adds a few new elements. Let us first map GePe to DAC:

- Grounded in the experiencing of physically and socially instantiated self: the somatic layer constitutes the foundation of the embodied hierarchy.
- Co-defined in the sensorimotor coupling of the agent to the world: both the reactive and adaptive layers establish immediate sensorimotor loops with the world (the former predefined, the latter acquired). Acquired sensorimotor states form the representational building blocks of DAC's cognitive processes.

- Maintained in the coherence between sensorimotor predictions of the agent and the dynamics of the interaction with the world: the adaptive level relies on prediction-based systems for both perceptual and behavioral learning (Duff and Verschure 2010). The memory systems of the contextual layer operate on a combination of forward and feedback models.
- DAC combines high levels of differentiation (each conscious scene is unique) with high levels of integration: the contextual layer integrates across all sensory modalities and memory systems and provides selection mechanisms to define a unique interpretation of the state of the world and the agent.
- Consciousness depends on highly parallel, distributed implicit factors with metastable, continuous, unified explicit factors: the contextual layer integrates memory-dependent implicit biases in decision making and interpretation of states of the world with explicit perceptual states. Task relevant states are "ignited" by the confluence of perceptual and memory evidence to form the dominant state of the contextual layer memory system.

If DAC resonates so well with GePe, why has it not reported conscious states? The answer is simple: GePe is incomplete. The DAC-based theory of consciousness, however, adds new considerations to the GePe framework:

1. *Simulation and its virtualization memory*: The hidden states of the world (i.e., other agents) are resolved through simulations which allow predictions on hidden states to be generated and maintained through forward models. As a result, action takes place in an augmented reality where sensor data (reflecting physical sources of stimulation and projected intentional states) are merged and tested against the world. This augmentation cannot take place in the physical world and thus requires a dedicated memory system which supports the virtualization of the world model. Some have characterized such a feature as a brain-based virtual reality (Revonsuo 1995; Merker 2005; Metzinger 2003).
2. *The intentionality prior*: To bootstrap the semantics of the simulations of hidden states of other agents, they are anchored in an intentionality prior, or pervasive intentionality, where novel states are automatically treated as being caused by other agents (Verschure 2012a). This implies that intentionality detection is operating at the level of the reactive layer. A further interpretation of intentional cues detected in the world or ascribed to it capitalizes on a self as other process (Merleau-Ponty and Edie 1964), which implies that the self and world columns of the architecture (see Figure 14.2) are tightly coupled, and that the self model continuously serves as an anchor of intentional cues detected in or projected onto the world.

3. *Parallel multi-scale operations*: Given the number of variables to be considered in a complex multiagent world and the finite operation powers of physical systems (i.e., brains), there is strong pressure on implementing components 1 and 2 through parallel operations. In addition, all real or imagined agents in the environment must be tracked in real time, thus defining a further functional need for parallelization. Indeed, parallel processing is one of the characterizing features of social brains, from the mushroom bodies of bees to the cerebellum of vertebrates.
4. *Serialization and unification*: The agent and its physical instantiation by necessity can only commit to a single action realized through its singular body at each point in time. These actions are informed by massively parallel simulations of possible world states that support real-time inference in an intention-laden world, thus creating a fundamental credit assignment problem: Given the outcome of a singular act, which value or action should be assigned to which property of the real or imagined world?
5. *Consciousness solves credit assignment in a parallel world model*: DAC allows us to rephrase the challenge of finding alignment between the singular and serial self model with a parallel and probabilistic world model. Real-time control of action requires parallel processing. For instance, the human cerebellum (credited with controlling real-time action) comprises about 15 million parallel segregated loops, constituting about 70% of the neuronal volume of the brain. Learning in this system is regulated through an error signal generated by the inferior olive, which matches reactive and adaptive modes of control (Herreros and Verschure 2013). Consciousness is a necessary counterpart to such a real-time parallel control system: a highly integrated sequential process that runs adjacent to the many parallel unconscious processes, integrating across many parallel states, valuating performance and projecting back error signals. In this way, cooperation between parallel unconscious and serial conscious control assures operational coherence through the reinterpretation and optimization of unconscious parallel loops. To realize this function, the process of consciousness requires a transient memory system that maintains the serialized and unified description of the world model in terms of the self-model. Unified intentionality is subsequently ascribed to the world and interpreted based on sequential conscious processing, in which self-generated actions are (re) interpreted, valued, and reorganized for future use (Verschure 2012b). Hence, consciousness serves goal-oriented performance in the future in a world filled with intentionality, while real-time action is under the control of the parallel unconscious systems that it optimizes (Figure 14.4). The problem of unifying the optimization of subconscious control is thus solved by shifting the representational frame from signal-based to intention-based, i.e., an intentional stance (Dennett 1988)

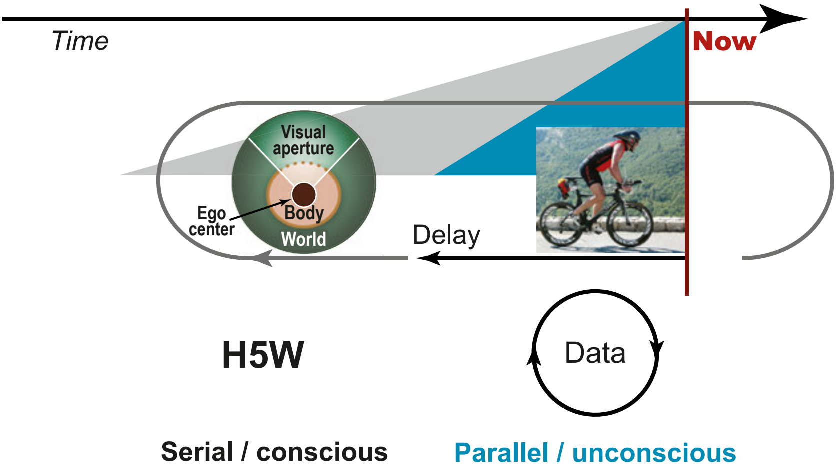

Figure 14.4 The interplay of parallel, real-time unconscious processing and sequential conscious monitoring. The green-filled circle represents the delayed serial conscious experience derived from the parallel control that generates real-time action.

framework, with respect to both the world and the self. The latter is achieved through "*ap*-presentation" (Merleau-Ponty and Edie 1964)—that is, the interpretation of the other in terms of the self—as well as by relying on the *a priori* ascription of intentionality to the environment, or the pervasive intentionality prior (Verschure 2012a). We can take this pervasive intentionality as another Kantian prior. Indeed, as Dan Sperber puts it: "the attribution of mental states is to humans what echolocation is to bats" (Gallagher 2005:207). Humans make social judgments based on simple geometric shapes (Heider 1944) or moving point models of the human body (Scholl 2001). Hence, the assumption of an intentionality prior as acting already at the level of reactive control seems to be defendable and a small price to pay to save consciousness from explanatory nihilism.

Addressing the "Hard" Problem: A Methodological Proposal

How can we make these predictions measurable or, more generally, how can we overcome the so-called hard problem? First, we do not need to ask physics to explain the form of a chair or even to provide a full explanation of the properties of materials. In fact, the dream of a unity of science is still stuck at crossing that bridge. Second, it can be argued that the hard problem is equally difficult in memory research as it is for consciousness: What is it like to remember some episode experienced in the past, and how is this information stored and retrieved? However, this might avoid essential issues, so let us consider a third option. The explanatory gap can be crossed when we focus on the *process* of consciousness rather than insisting that each specific quale be deciphered. DAC theory has progressed by insisting on a methodology of convergent validation (Verschure 1997): constraining models through simultaneously addressing anatomy, physiology, and behavior. As a result, all DAC models are validated using real-world behaving systems (i.e., robots). This approach is grounded in the philosophy of the eighteenth century Neapolitan philosopher Giambattista Vico, who famously proposed that we can only understand that which we create: *Verum et factum reciprocantur seu convertuntur* (Vico 1730). We can use this same approach to address the hard problem and parse quale, which are essentially expressing memory and which, in turn, reflect a specific prior experience at a specific point in time by a specific agent with its specific embodiment and history. We merely have to follow the machine in time and record its states to be able to parse its mental states, including conscious ones, on future occasions. We have already performed such semantic parsing of memory states successfully with complex autonomous DAC-based artifacts, such as the interactive sentient space Ada visited by over 500,000 humans (Eng et al. 2005) and different behaving robots (Verschure et al. 2003). Hence, testing a theory of consciousness requires us to build a conscious machine and there is no *a priori* reason why this should be impossible.

Empirical Consequences

The H5W DAC theory of consciousness provides an explanation of consciousness that focuses on the notion of the unitary nature of conscious experience and its delayed realization relative to real-time performance. The model behind it is advanced through controlling real-world systems, including buildings and robots. To complete all criteria of a scientific theory, the question is: What testable predictions can we derive from H5W DAC?

The first candidate is the assumption of pervasive intentionality or the intentionality prior. This suggests that infants would ascribe excessive intentionality to the world and that as a result of maturation, this intention reflex is suppressed. Indeed, it has been shown that both seven-month-old infants and adults model the intentional states of others in a form similar to their self models (Kovács et al. 2010). In addition, there is a negative correlation between age and the propensity to favor teleological explanations of social behavior, biological properties, artifacts, and life events (e.g., Banerjee and Bloom 2015). The fact that infants, starting at five months of age, show slowly maturing neurophysiological signatures of consciousness perception (Kouider et al. 2013), opens up a wide range of experimental questions on the relation between consciousness, pervasive intentionality, and maturation (e.g., how 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.

when the intentional reflex is suppressed and how this is reflected in the signatures of conscious experience).

We have looked directly at the question of the hierarchical structuring of conscious experience in an effort to disentangle subconscious from conscious processing in the context of goal-oriented psychophysical tasks (Mathews et al. 2015). Using a displacement detection task combined with reverse correlation, it was shown that bottom-up fast saccades, top-down driven slow saccades, and conscious decisions follow distinct regions of the sensory space, or validation gates, modulated by the conscious task the subject performs. This experiment demonstrated that conscious decision making can be largely dissociated from subconscious parallel processing; it also provides support for the DAC notion of parallel layered control and for the view that consciousness provides a time-delayed description of an effective task that comprises a subset of the world in which the subject is acting. In addition, the idea of a continuous perceptual hierarchy linking sensation to perception and experience, popular in current Bayesian brain accounts (see Friston, this volume) does not hold up under these conditions; such hierarchical relations are dynamically formed dependent on the task conditions faced by the self.

Conclusion

In this chapter, I have presented the DAC H5W theory of consciousness in the context of a historical cycle in the study of mind, brain, and behavior and discussed the GePe framework, capturing contemporary science and philosophy of consciousness studies. I propose that returning to the question of the function of consciousness and its implementation by the brain is a historical prerogative, if we want to avoid sliding down the sinkhole of big data and the scientism of explanatory nihilism. It is important to emphasize that we should avoid the fallacy of mistaking our measures for the phenomena they are designed to measure, as behaviorism discovered at its peril. One can interpret the move into neo-panpsychism of the proponents of IIT as an artifact of such a scientism fallacy. Indeed, IIT and the Bayesian brain frameworks illustrate a trend to collapse the complexity of mind and brain into relatively simple quantitative measures, uncoupled in any relevant way to the neuronal substrate or action (i.e., the levels of observation that are accessible for a science of mind). The global workspace framework faces a similar problem of multi-instantiation; it could be realized in arbitrary hardware systems and is not constrained by any fundamental property of the brain. Hence, the challenge is to derive a convergent science of consciousness that is able to show how the brain generates and expresses consciousness in action.

I propose that consciousness is critically related to action in an intentional world or the transition from an agent that solves H4W to solving H5W.

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.

Consciousness provides the interface between the singular self and the parallel world. In this proposal, conscious is by necessity intentional because it pertains to a single agent engaged with an intentional world. Grounded in the physical existence of the agent over time, the self-constructed conscious narrative defines its subjectivity and quale and assures the coherence of its operation. Thus consciousness is the coherent experience that results from the large-scale integration of perception, affect, memory, cognition, and action along the neuroaxis in a dedicated memory system. It is a form of memory that unifies and interprets the states of the agent to facilitate the optimization of its parallel real-time control loops that are driving action. This memory is only active when the agent is. This H5W hypothesis predicts that the conscious scene is a transient memory implemented in the thalamocortical system, which provides a unitary description and valuation of real-time performance and is able to project this valuation onto the parallel control loops of the brain (e.g., those found in the cerebellum). In this way the solution to the H5W problem reinstates free will as the ability to *will* improvement of performance in the future, as opposed to stopping at the contemporary interpretation that we lack the will to control our performance in the "now." Thus, the H5W hypothesis aims at explaining consciousness as a natural phenomenon—a property of specific biological systems that emerged during the Cambrian to act in an intentional world to survive. It would be premature to say that DAC has *explained* consciousness, but we can observe that it does capture the main components of the GePe framework, while advancing a concrete research agenda that poses specific questions about perception, emotion, cognition, and actions structured along H5W. With this in hand, we can turn to the more specific question of the functional role of consciousness and what this would imply for future extensions of the DAC theory.

Others have also advanced hypotheses which emphasize the social origins of consciousness (Mead 1934; Humphrey 2006; Baumeister and Masicampo 2010; Graziano 2013). These proposals have emphasized the contribution of consciousness to specific aspects of social interaction (e.g., rational thought, language, attention). The DAC H5W hypothesis emphasizes the role of consciousness in optimizing the control structures that social interaction and its underlying intentional stance requires.

As we continue to analyze the pragmatic turn in cognitive science, we need to be mindful of the damage previous forms of pragmatism have caused: behaviorism was the primary cause behind the disappearance of consciousness from the scientific landscape, because its methods and philosophy could not address it. A similarly dogmatic narrow view must be avoided at all costs. We need to return to a science that insists on gnawing the file of consciousness until it gives way to a deeper understanding of nature and ourselves.

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.

Acknowledgments

I am grateful for the feedback I have received to an earlier draft of this chapter by the two reviewers appointed by the Forum. The work reported here is supported by the EU Integrated Project CEEDS funded under the Seventh Framework Programme (ICT-258749) and ERC grant cDAC (ERC-341196).

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.