HISTORICAL PERSPECTIVES ON THE WHAT AND WHERE OF COGNITION

Lena Kästner and Sven Walter

University College London
Institute of Cognitive Neuroscience
17 Queen Square
WC1N 3AR
London
United Kingdom

mail@lenakaestner.de

University of Osnabrück
Institute of Cognitive Science
Albrechtstraße 28
D-49069 Osnabrück
Germany

s.walter@philosophy-online.de

What is cognition? The embarrassing answer is: There is no unanimously accepted answer, not even remotely. We simply don't seem to know.

We don't know *yet*, optimists insist. We have a list of fairly uncontroversial prototypes of cognitive processes—categorization, learning, perception, reasoning etc.—and it is only a matter of time until we can specify a "mark of the cognitive" capturing their common core. The problem is: There is no reason for thinking the mechanisms implementing these prototypes will have anything significant in common, materially or functionally.

We will *never* know, pessimists insist. "Cognition" is merely a label for a motley bundle of processes that are of interest to cognitive scientists for some reason or other. Many sciences invoke key concepts that lack crisp and clear definitions—"language" in linguistics, say, "gene" in biology, or "intelligence" in psychology. "Cognition" may just be another case where researchers recognize a phenomenon if they come across it, but are unable to provide necessary and sufficient conditions.

But this agnosticism cannot succeed. We are interested not only in the *What*, but also in the *Where* of cognition. Clark & Chalmers (1998) famously maintained that the material vehicles of some cognitive processes *extend* beyond the brain and the body into the environment. Defenders of a more conservative view, in contrast, insist that cognition resides in brains, or at least bodies. The only sensible way of settling this dispute is to provide a mark of the cognitive and then go and see where in the world the processes fulfilling it are found. Resolving the *Where*-question thus presupposes resolving the *What*-question. This is why agnosticism is untenable—we must know *what* cognition is before we can make any progress on its *where*.

Agnosticists may try to dismiss the Where-question, arguing that extended cognition is merely a fancy philosophical hypothesis. A look at the history of cognitive science shows that this is not true. The idea of cognitive extension is a natural consequence of earlier approaches to cognition, and it is a legitimate question to ask whether it is correct or not. A look at the history also allows us to understand the importance of, and the various answers to, the What- and Where-questions.

1. Classicism

Scientific work on computation, information theory and cybernetics during and after World War II culminated in the interest in the artificial design of intelligent agents that gave AI its name. Two "classicist" views of cognition originated from this early work: the more theoretically inspired algorithmic "rules and representations" approach of *good old-fashioned artificial intelligence* (GOFAI) and the more biologically inspired neural network approach of *connectionism*.

GOFAI's answer to the What-question was: cognition is algorithmic *information-processing* in the sense of rule-governed, sequential *computations* over structured symbolic *representations*. Since in humans these representations are arguably encoded neurally, the computational processes in question are an entirely intracranial affair. GOFAI's answer to the Where-question thus was: in the head. The world serves only as a source of perceptual input and the arena for behavioral output, while all the cognitive processing is done in the head; cognition is a central element "sandwiched" (Hurley 1998) between the peripheral buffer zones of perception and action.

According to connectionism, cognition is grounded in spreading activation in heavily connected networks of neuron-like information processing units. *Information-processing* thus again played a crucial role, and so did *computation*. Computations in connectionist networks, however, are not rule-based (not explicitly, at least), they are *local*, i.e., they take place at the level of individual network nodes, and thus *parallel* in the sense that multiple nodes are simultaneously active: A single node is usually involved in a range of a network's states, and its current activity is determined by the network's overall activation pattern. Although this starkly contrasts with GOFAI's conception of computation as a rule-based, global, sequential process, it is computation nevertheless. *Representations* also remained a crucial element. But since individual nodes do not normally map in a one-to-one fashion onto the constituents of what the network's overall state stands for, they were said to be *subsymbolic* or *distributed* representations. Hence, although the details were different, connectionism's answer to the

What-question was essentially the one already given by GOFAI—information-processing by computations over representations. Connectionism obviously also endorsed the sandwich model. Moreover, since the relevant networks in humans are their brains, the answer to the Where-question was the same, too: cognition is an entirely intracranial affair.

2. Dynamicism

Classicism was most successful in modeling disembodied, abstract features of human cognition that can be performed "off-line," i.e., detached from the world—like inference drawing and problem solving (GOFAI) or pattern recognition (connectionism). In contrast, advocates of a dynamicist approach emphasized the importance of "on-line" cognition: cases where cognitive systems are dynamically coupled to their environments in immediate, real-time interactions, and under continuous reciprocal causal influence (the distinction between "off-line" and "on-line" cognition is Wheeler & Clark's (1999)). Dynamicists stressed that brains are seamlessly integrated into their bodily and extrabodily environments in such a way that neurophysiological, physiological, and environmental processes form a single, dynamically changing whole. We should therefore treat cognitive systems as *dynamical systems* and model them by sets of differential equations: "cognitive agents are dynamical systems and can be scientifically understood as such" (Van Gelder 1999, 13). As a consequence, dynamicists downplayed the role of *computation*: "Rather than computers, cognitive systems may be dynamical systems; rather than computation, cognitive processes may be state-space evolution within these very different kinds of systems" (Van Gelder 1995, 346). They also eschewed the appeal to *representations*: since dynamical systems are continuously evolving, there are no discrete, sequential steps in which one representation is transformed into another (Van Gelder 1995): "We are not building representations of the world by connecting temporally contingent ideas. We are not building representations at all! Mind is activity in time …" (Thelen & Smith 1994, 338). Yet, mathematically speaking, dynamical systems are characterized by sets of state variables and sets of laws determining how the values of these variables change over time. Each possible state of a system is a point in its state space and a sequence of states is a trajectory through that state space. One may thus argue that, e.g., these trajectories through state space are, albeit in a weak sense, representations of the system's behavior.

The dynamicist's official view on the What-question is: cognition is state-space evolution in a dynamical system and thus neither decidedly computational nor decidedly representational, although still fundamentally information processing. The sandwich model is rejected, for the world is more than merely the passive source of input for and receiver of output from the cognitive system: "The cognitive system does not interact with other aspects of the world by passing messages or commands; rather, it continuously coevolves with them" (Van Gelder & Port 1995*a*, 2). The dynamicist may also be read as offering a radical answer to the Where-question. Since cognitive systems are a dynamically changing whole comprising brain, body, and environment, the skull no longer looks like a natural boundary for the cognitive: "Cognitive processes span the brain, the body, and the environment" (Van Gelder & Port 1995*b*, ix). Understood that way, dynamicism seems to support the idea of *extended cognition*. However, as dynamical systems are abundant, cognitive processes can at best be a *subset* of dynamical processes; and unless the dynamicist specifies what exactly it is that makes a dynamical process cognitive, her answer to the Where-question may be interpreted conservatively: although dynamical processes are everywhere and crisscrossing the boundaries of brain, body, and environment, the dynamical processes that are cognitive may all reside in the brain.

3. Situated Cognition

Dynamicists criticized the insular, sandwiched view of cognition characteristic of classicism and instead stressed the importance of *body* and *environment*. Like dynamicism, situated approaches to cognition argued that classicism focused too narrowly on abstract programs for specialized feats of reasoning and inference in highly specialized domains, thereby neglecting that cognition emerges "on-line" out of the interactions between embodied cognitive systems and their environments, rather than being done "off-line" by a detached computational and representational system implemented in the brain. Understanding cognition thus requires understanding how physically embodied agents achieve sensorimotor control in "fluid and flexible" (Wheeler 2005, 170) real-time interactions with their environment. The slogan was to put cognition "back in the brain, the brain back in the body, and the body back in the world" (Wheeler 2005, 11). The resulting situated approaches to cognition are a relatively recent development with a variety of subtly different strands whose key tenets, theoretical and terminological commitments, and interrelationships and interdependencies are still in disarray (Robbins & Aydede 2009). Below we offer a taxonomy that strikes us as plausible (Walter 2010*a*), but others may disagree about the correct classification of the various approaches.

3.1 Embodied Cognition

According to the embodied approach, cognition bears a profound relation to bodily processes in the sense that "the specific details of human embodiment make a special and … ineliminable contribution to our mental states and properties" (Clark 2008*b*, 39). Pioneering work in the embodiment paradigm came from Rodney Brooks' bottom-up robotics (Brooks 1991), but quite generally the embodied approach to cognition is the attempt to carry out what Anderson (2003) called the "physical grounding project," viz. to show exactly how an agent's physical features and abilities contribute to her cognitive processing. Within the embodied cognition paradigm, the grounding relation is spelled out in at least two different ways.

Paradigm examples of the embodied approach to cognition are Lakoff & Johnson's (1999) work on our body's contribution to our conceptual repertoire and McBeath *et al*.'s (1995) study on how baseball outfielders manage to catch fly balls. According to Lakoff and Johnson, all our concepts are ultimately derived from basic concepts that stem directly from and are constrained by the type of body we possess (e.g., spatial ones like *up*, *down*, *front*, *back* etc.). According to McBeath *et al*., a classicist sandwich solution to the problem of catching a fly ball would be to take the visual perception of the ball as input, generate an internal representation, let an internal reasoning system use that representation to compute the ball's future trajectory, and finally trigger an appropriate motor output. In reality, McBeath *et al*. argue, the solution relies on certain characteristics of the outfielder's body, thereby minimizing the need for internal computation and representation: simply run in such a way that the optical image of the ball appears to present a straight-line constant speed trajectory against the visual background.

Both examples illustrate how "the presence of a humanlike mind *depends* quite directly upon the possession of a humanlike body" (Clark 2008*b*, 43; emphasis added). The embodied approach therefore rejects the sandwich model. A straightforward positive answer to the What-question, however, is lacking. On the one hand, as the research of McBeath *et al*. illustrates, there is a certain depreciation of the role of computation and representation (also highlighted in Brooks' work on bottom-up architectures in robotics), but on the other hand Lakoff and Johnson's work is rather neutral with regard to these issues. No clear answer to the What-question, thus. The answer to the Where-question, in contrast, is clear, and it is still conservative: Cognitive processes, although *causally dependent* upon extracranial bodily processes, are an entirely intracranial affair.

According to a stronger version of the embodied approach, cognitive processes are not only *dependent* upon but actually *constituted* by bodily processes. Support for this stronger claim comes from studies showing that vision essentially relies on bodily movements (e.g., Ballard *et al*. 1997; Noë 2004; O'Regan 1992; O'Regan & Noë 2001). Shapiro argues that bodily movements are not only extracranial aids but "as much *part* of vision as the detection of disparity or the calculation of shape from shading" (2004, 188), so that "[v]ision for human beings is a process that *includes* features of the human body" (2004, 190; both emphases added). As in the case of the weak embodied approach, the implications with regard to the role of computation and representation are unclear. While Noë (2004) is usually pictured as a strict anti-representationalist, Ballard *et al*. are clearly in favor of representations and argue that their model "strongly suggests a functional view of visual computation" (1997, 735) able to combine the idea that vision is a form of acting with the idea that it is a computational process. The strong embodied approach gives a more radical answer to the Where-question, however: Cognitive processes include *extracranial bodily* processes and are thus not merely in the head. This ambivalence in the embodied approach is rarely noted in the literature, although it has obvious ramifications for the proper study of cognitive processing: According to the weaker version, cognitive processes are restricted to an organism's brain, while according to the stronger, they are leaking out into the organism's body.

3.2 Embedded Cognition

Embedded approaches to cognition stress the role of the environment and its active structuring by the agent. Recent research on visual processing (Noë 2004; O'Regan & Noë 2001) suggests that instead of creating detailed internal representations as the basis for later stage cognitive processing, human subjects extract the relevant information "on the fly" from the world itself. Kirsh & Maglio's (1994) research on "epistemic actions" highlights a similar kind of environmental "offloading" or "outsourcing" of cognitive load. Experienced *Tetris* players rotate the figures on the screen rather than mentally because it is cognitively less demanding. The important point is not just that the way the world is influences an agent's cognitive processing; it is that the agent herself *actively* structures her environment in order to facilitate cognitive processing. The embedded approach may presuppose the embodied approach in the sense that an agent's capacity for cognitive off-loading depends not only on the environment but also on her body, because it is her body which determines how she can perceive, navigate and manipulate her surroundings.

Regarding a positive answer to the What-question, the embedded approach is—again—not particularly forthcoming. What is clear is that while the *computational* nature of cognition is typically not denied, the idea of off-loading shows that the computations in question may directly involve extrabodily items rather than their rich internal *representations*: Why bother representing something internally that is right there in your environment? Simply use, as Brooks famously put it, the world itself as "its own best model" (Brooks 1991, 583). The embedded approach is decidedly more radical than both embodied approaches, for it entails that cognitive processes must be studied not by looking at their causal (weak) or constitutive (strong) grounding in extracranial bodily processes, but by looking at the way an agent uses her environment's structure or actively structures her environment. Regarding the Where-question, the embedded approach is a natural extension of the weak embodied approach. Like the weak embodied approach, it specifies the grounding relation in terms of *causal dependence*; unlike the weak embodied approach, however, it takes the dependence base of cognitive processes to contain not only extra*cranial bodily*, but also extra*bodily* processes. Cognitive processing thus takes place in the brain and in the extracranial parts of the body, although it causally depends upon the extrabodily environment.

3.3 Extended Cognition

From embodied and embedded cognition it is only a short step to extended cognition. If body and environment are indeed crucial for cognitive processing, then they may literally be a *part* of—rather than merely causally contributing to—cognition. Just as the embedded approach extends the dependence base of the weak embodied approach from extracranial bodily processes to extrabodily processes, extended cognition is a natural corollary of the strong embodied approach. Like the strong embodied approach, it stresses that cognitive processes are partially *constituted* by extracranial processes; unlike the strong embodied approach, however, the constituents of cognitive processes are taken to be not only extra*cranial*, but also extra*bodily*.

Regarding the What-question, most advocates of the extended approach would apparently be prepared to endorse the claim that cognition is a computational information-processing process, albeit one which consists in computations over internal or external representations or even the extrabodily items themselves: "at least some of the computational systems that drive cognition reach beyond the limits of the organismic boundary" (Wilson 2004, 165). The extended approach is thus surprisingly conservative. All it does is to allow for computational processes to range not only over internal representations, as in classicism, but also over external representations or the extrabodily items themselves. Unlike classicism, however, the extended approach rejects the sandwich model, for the world is an active part of cognition, not only the passive source of input and the stage for output. With regard to the Where-question, the extended approach is the most liberal: Cognitive processing involves intracranial, extracranial bodily and extrabodily processes.

4. The What of Cognition, Again

This admittedly brief overview shows that the extended approach is not merely a fancy philosophical idea with no basis in cognitive scientific practice. There are two routes to the idea of cognitive extension, one *via* dynamicism, and one *via* embodied and embedded approaches. The *Where* of cognition is thus a substantial and important issue that needs to be resolved. As said in the beginning, agnosticism is not a viable option because answering the Where-question arguably requires answering the What-question (Walter 2010*b*). Unfortunately, as the preceding considerations have shown, there is not even a remotely unanimously accepted answer to the question "What is cognition?," except for the idea that cognition probably has something to do with information-processing, which can at best be a necessary, not a sufficient condition. Table 1 summarizes the results.

INSERT TABLE 1 AROUND HERE

Thus far, it seems, we do not know what is distinctive about cognition, nor do we understand how it works, or where to look for it. The situation seems rather bleak. What are the options? Rather than immediately delving into the details of specific accounts of cognition, we suggest that it may be worthwhile to first ask a more general question: Should "cognition" be taken to be a *natural kind term*, a *cluster term*, or an *umbrella term*?

4.1 "Cognition" as a Natural Kind Term

The perhaps most intuitive approach to answering the What-question is to assume that "cognition" is a natural kind term whose instances have a scientifically discoverable essence—what Adams & Aizawa (2008, 2009) famously call a "mark of the cognitive." Cognitive processes, they argue, "are natural kinds of processes" (2008, 80) consisting in computational operations that involve *non-derived representations* and are implemented by special kinds of *mechanisms*. Since non-derived representations and the kinds of mechanisms at issue are found, currently at least, only in the brain, there is "defeasible reason to suppose that cognitive processes are typically brain bound and do not extend from the nervous system into the body and the environment" (2008, 70). Adams and Aizawa's natural kind conception of the cognitive thus entails a conservative answer to the What-question: Cognitive processes are, as a contingent matter of fact, found in the head and only in the head.

Although this is not the place to go into the details, note that a number of complaints can be leveled against Adams and Aizawa's approach. First, since there is no received theory of non-derived content, we cannot tell whether a process fulfills their mark or not. Second, unless there is a theory of non-derived content, it is hard to substantiate their claim that non-derived representations are currently found in the brain alone and not in the brain *cum* body *cum* environment. Third, Adams and Aizawa's claim that cognitive mechanisms must be individuated in terms of their material implementation begs the question against the functionalistic approach usually adopted by defenders of an extended approach (Walter 2010*b*). Finally, prematurely equating cognition with specific kinds of brain processes forecloses fruitful future discoveries in cognitive science.

Despite this skepticism about the approach of Adams and Aizawa, they do seem to have a point. Any conception of the cognitive that could support an extended view would arguably have to cover so heterogeneous processes that "cognition" would fail to pick out a natural kind. In other words: If "cognition" is a natural kind term, the answer to the Where-question is most likely going to be conservative.

4.2 "Cognition" as a Cluster Term

Given the broad range of phenomena we usually count as cognitive and given our apparent difficulties in capturing a common essence, "cognition" may simply fail to pick out a natural kind. There may just not be a set of individually necessary and jointly sufficient conditions for a process' being cognitive. Since "cognition" could be a *cluster term,* any cognitive process could still share some of its characteristics with other cognitive processes, but they would only amount to a *family resemblance*. Wheeler (2005), for instance, suggests that cognitive processes can be implemented by (i) non-computational and non-representational, (ii) non-computational and representational, and (iii) computational and representational mechanisms. In that case, finding a "mark of the cognitive" would mean identifying the structure of the underlying resemblances rather than a single common core that fixes the meaning of the term "cognition." Such an approach would seem to be compatible with all sorts of answers to the Where-question, depending upon where in the world the processes in question are found.

4.3 "Cognition" as an Umbrella Term

A third way of thinking about the demarcation of the cognitive would be to loosely characterize the basic commonalities of individual cognitive phenomena while accepting that they neither form an overarching natural kind of "the cognitive" nor exhibit any family resemblances. "Cognition" would then be an *umbrella term* under which, as Clark proposes, a "motley crew of mechanisms" (Clark 2008*a*) finds shelter. The concept of "memory" may illustrate the idea. It was only after discovering dissociations in neuropsychological patients like H.M. that the different mechanisms underlying what formerly seemed to pick out a single general capacity for memory had been recognized. Subsequent research revealed that memory decomposes into a variety of phenomena, the most general distinction being between *long-term memory* (LTM) and *short-term memory* (STM), both of which allow for further subdivisions. Analogously, cognition may decompose into a range of causally distinct processes "with not even a family resemblance" (Clark 2008*a*, 95).

Decomposing cognition into subtypes may seem fairly reasonable once we reconsider the broad range of phenomena cognitive scientists are interested in. The important open question, however, is what unifies the umbrella's subtypes if there are not even family resemblances. Clark (2008*a*) suggests that cognition should best be understood as *information processing tightly coupled* to a *cognitive core*—probably, but not necessarily, the brain. However, Clark's characterization of the cognitive requires an account of "information," "coupling," and "cognitive core," and it is hard to see how to spell out the notion of a "*cognitive* core" without having already at hand a notion of the cognitive. Most importantly, however, Clark's approach is an instance of what we have dubbed "agnosticism." And as we have indicated in the beginning, it is a mistake to think that the Where-question can be successfully answered against the background of an agnostic stance towards the What-question.

5. Conclusion

We started with the claim that there is at least one good reason to ask the What-question: In order to settle the Where, we *have to* answer the What-question. Once we have done that, we can simply go and look where in the world we find cognition.

This means we need a mark of the cognitive. The prospects for such a mark, however, depend on what sort of term "cognition" is, as section 4 has shown. Moreover, since each of the paradigmatic options presented here struggles with shortcomings, the situation seems rather bleak. Being aware of the options available to us, however, may help to guide cognitive scientific research and eventually enable us to answer both the What- and the Where-question.

References

Adams, F.–Aizawa, K. (2008): *The Bounds of Cognition*. Oxford: Balckwell.

Adams, F.–Aizawa, K. (2009): Why the mind is still in the head. In: Robbins, P.–Aydede, M. (eds.), *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press, 78–95.

Anderson, M. (2003): Embodied cognition: A field guide. *Artificial Intelligence* 149, 91–130.

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

Brooks, R. (1991): Intelligence without reason. *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, 569–595.

Clark, A. (2008*a*): *Supersizing the Mind*. New York: Oxford University Press.

Clark, A. (2008*b*): Pressing the flesh: A tension in the study of the embodied, embedded mind? *Philosophy and Phenomenological Research* 76, 37–59.

Clark, A.–Chalmers, D. (1998): The extended mind. *Analysis* 58, 7–19.

Hurley, S. (1998): *Consciousness in Action*. Cambridge, MA: Harvard University Press.

Kirsh, D.–Maglio, P. (1994): On distinguishing epistemic from pragmatic action. *Cognitive Science* 18, 513–549.

Lakoff, G.–Johnson, M. (1999): *Philosophy in the Flesh*. New York, NY: Basic Books.

McBeath, M.–Shaffer, D.–Kaiser, M. (1995): How baseball outfielders determine where to run to catch fly balls. *Science* 268, 569–573.

Noë, A. (2004): *Action in perception*. Cambridge, MA: MIT Press.

O’Regan, K. (1992): Solving the “real” mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology* 46, 461–488.

O’Regan, K.–Noë, A. (2001): A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24, 939–960.

Robbins, P.–Aydede, M. (2009): A short primer on situated cognition. In: Robbins, P.–Aydede, M. (eds.), *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press, 3–11.

Shapiro, L. (2004): *The Mind Incarnate*. Cambridge, MA: MIT Press.

Thelen, E.–Smith, L. (1994): *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Van Gelder, T. (1995): What might cognition be if not computation? *Journal of Philosophy* 92, 345–381.

Van Gelder, T. (1999): Defending the dynamical hypothesis. In: Tschacher, W.–J. Dauwalder (eds.), *Dynamics, Synergetics, Autonomous Agents*. Singapore: World Scientific, 13–28.

Van Gelder, T.–Port, R. (1995*a*): It's about time: A perspective to dynamical system approach to cognition. In: Port, R.–Van Gelder, T. (Eds.), *Mind as Motion*. Cambridge, MA: MIT Press, 1–43.

Van Gelder, T.–Port, R. (1995*b*): Preface. In: Port, R.–Van Gelder, T. (Eds.), *Mind as Motion*. Cambridge, MA: MIT Press, vii–x.

Walter, S. (2010*a*): Locked-in syndrome, BCI, and a confusion about embodied, embedded, extended, and enacted cognition. *Neuroethics*.

Walter, S. (2010*b*): Cognitive extension: The parity argument, functionalism, and the mark of the cognitive. *Australasian Journal of Philosophy*.

Wheeler, M. (2005): *Reconstructing the Cognitive World*. Cambridge, MA: MIT Press.

Wheeler, M. (MS): *Extended X*. http://www.philosophy.stir.ac.uk/staff/m-wheeler/ExtendedX.php; accessed October 20, 2009.

Wheeler, M–Clark, A. (1999): Genic representation: Reconciling content and causal complexity. *British Journal for the Philosophy of Science* 50, 103–135.

Wilson, R. (2004): *Boundaries of the Mind*. Cambridge: Cambridge University Press.