EXTENDED COGNITION: FEEDBACK LOOPS AND COUPLED SYSTEMS

Olga Markič*

University of Ljubljana – Faculty of Arts, Department of Philosophy
Ljubljana, Slovenia

DOI: 10.7906/indecs.15.4.5

Regular article

Received: 14th December, 2017.
Accepted: 22nd December, 2017.

ABSTRACT
The article explores two waves of active externalism. I first introduce the distinction between passive and active externalism and analyse a proposal of active externalism based on the principle of parity proposed by Clark and Chalmers. There are two main obstacles, causal-constitution fallacy and cognitive bloat, that threaten the extended cognition hypothesis. The second wave of discussions based on the complementarity principle deals with cognitive systems with feedback loops between internal and external elements and is a more radical departure from functionalism and traditional thinking about cognition. I conclude with some remarks on potential ethical considerations of extended cognition.

KEY WORDS
extended mind, extended cognition, principle of complementarity, coupled systems

CLASSIFICATION
APA: 2340, 2380, 2630
JEL: Z13

*Corresponding author, η: olga.markic@ff.uni lj.si; +386 1 241 1104;
Faculty of Arts, Aškerčeva 2, SI-1000 Ljubljana, Slovenia
INTRODUCTION

Andy Clark and David Chalmers begin their famous article *The Extended mind* with the following question: “Where does the mind stop and the rest of the world begin?” [1; p.27]. Although it may seem that finding the right answer should be quite straightforward, it turns out that it is not. First, it tackles one of the basic problems in philosophy of mind, the mind-body problem. For example, if you are a dualist and follow the philosopher René Descartes [2], you will put the mind as a separate substance, *res cogitans*. According to his view the mind interacts with the body in the human brain, more specifically in the pineal gland and thus form a mind-body unity. The interactionist dualistic position has been criticized ever since it was proposed, mainly because it seems incompatible with requirements of natural science [3], although it has also contemporary advocates [4]. In this article we will leave aside dualism/monism debates in philosophy of mind and assume a naturalist position that rejects supernatural entities and takes mind and cognitive processing as natural phenomena. But, by accepting naturalism we open a plethora of further question. For example, is mind identical with brain and consciousness is a brain process [5] or does mental emerges from complex processes in the brain [6]? The main focus of this article will be on the distinction between internalism and externalism, particularly on the question if externalism (and which version) provides better approach to study cognition.

I will first briefly sketch Clark and Chalmers’ idea of the extended mind and extended cognition as proposed in their article [1]. I will then introduce the distinction between passive and active externalism and analyze the proposal of active externalism based on the principle of parity. I will argue that the difficulties such an approach faces, namely causal-constitution fallacy and cognitive bloat, threaten the plausibility of the extended cognition hypothesis. I will move to the second wave of discussions based on the complementarity principle that presents a more radical departure from functionalism. I will suggest that systems with feedback loops between internal and external elements provide a promising approach to how extended cognition can escape the before mentioned difficulties. I will conclude with some remarks on potential ethical considerations of extended cognition.

PASSIVE AND ACTIVE EXTERNALISM

The idea that the mind is not just in the head but can be extended to the world has forerunners in semantic externalism and externalism about mental content. Externalism with regard to mental content is the position that our contents depend in a constitutive manner on items in the external world, both natural and social world. So, “in order to have certain types of intentional mental states (e.g. beliefs), it is necessary to be related to the environment in the right way” [7]. Internalism, on the other hand, is the position that denies this, our contents depend solely on our intrinsic properties, on properties of our bodies, such as our brains [7, 8]. The view of externalism is nicely summarized by one of the main proponents of semantic externalism, Hilary Putnam: “the content of sentences (and, derivatively, the content of beliefs and other language-dependent psychological conditions) is at least partly dependent on the determination of the reference in the particular context (in technical jargon, on the ‘extension’) of the terms used in the sentence or in the expression of belief, and that reference depends on factors that are external to the speaker’s body and brain. Whether, for example, a speaker means elm when she uses the word elm depends, inter alia, on whether her word refers to elm trees, and that depends in complex ways on both her relations to other speakers (in case the speaker, like so many of us, is unable to identify elm trees reliably on her own) and on what sort of trees are in fact in the environment of the speaker and of the experts on
whom the speaker relies. The speaker’s neurological condition (or ‘brain state’) may not in principle suffice to determine whether a given speaker refers to elm trees or to beech trees when she uses the word elm.” [9; p.119].

The long debate about semantic externalism has started with Putnam [10] and Burge [11]. Putnam presented a thought experiment about the Twin Earth. He imagines that somewhere there is a Twin Earth. People living on Earth and Twin Earth are exact physical duplicates and have the same behavioral histories, but there is one difference. The substance that we call water, on Twin Earth does not consist of H₂O, but of XYZ. He concludes that because natural kind terms like water refer to their hidden structure (to H₂O in the Earth and to XYZ in the Twin Earth), twins across planets, while they are in the same psychological state, mean different things when they say water. He thus concludes “Meaning just ain’t in the head”[10; p.227].

Clark and Chalmers [1] suggest that one has to go further than semantic externalism and the thesis that meaning can not be reduced onto internal states. Here is their way of reasoning: “When I believe that water is wet and my twin believes that twin water is wet, the external features responsible for the difference in our beliefs are distal and historical, at the end of a lengthy causal chain. Features of the present are not relevant: if I happen to be surrounded by XYZ right now (maybe I have teleported to twin earth), my beliefs still concern standard water, because of my history. In these cases, the relevant external features are passive. Because of their distal nature, they play no role in driving the cognitive processes in the here-and-now. This is reflected by the fact that the actions performed by me and my twin are physically indistinguishable, despite our external differences.” [1: p.29] In contrast to this weaker variant, called passive externalism, they advocate a stronger form they call active externalism, where “the relevant external features are active, playing a crucial role in the here-and-now” [1; p.29]. These relevant external features are coupled with the human organism and have a direct impact on the organism and on its behavior. They stress that “In these cases the relevant parts of the world are in the loop, not dangling at the other end of a long causal chain” [1; p.29]. Active externalism is thus more than merely causal thesis where external features in interaction with the organism causally influence cognitive processes. Clark and Chalmers point out that even if one accepts Putnam’s and Burge’s proposal about semantic externalism, it is not clear how external aspects play a causal or explanatory role in the generation of action. In the counterfactual cases when internal structure is held constant and only external features are changed, the behavior looks just the same and it seems that internal structure is doing the work. In contrast, active externalism they propose is not threatened by such problems, because “[t]he external features in a coupled system play an ineliminable role – if we retain internal structure but change the external features, behavior may change completely” [1; p.30]. According to their approach the external features are “just as causally relevant as typical internal features of the brain ” [1; p.30].

THE EXTENDED MIND AND THE EXTENDED COGNITION HYPOTHESES

There are two related formulations of active externalism within contemporary philosophy of mind: the extended mind and the extended cognition (HEC) hypotheses. According to the latter the cognitive processing can literally extend to the agent’s environment and features of the environment (e.g. pen and paper) are proper parts of the ongoing cognitive process. The extended mind thesis, instead of concentrating on cognitive processes, claims that mental states like experiences, beliefs and emotions get extended too. We can take this two hypotheses to differ in degree of radicalism [12]. I will start with the thought experiment that supports the extended mind, although I will later focus more on weaker HEC. In their thought experiment
Clark and Chalmers argue that beliefs can be constituted partly by features of the environment, “when those features play the right sort of role in driving cognitive processes” [1; p.33]. They introduce two persons, Inge and Otto, who are forming their beliefs about the Museum of Modern Art, each in her/his own way. Inge represents a normal case of a belief embedded in memory. She has heard from a friend that there is an exhibition at the Museum of Modern Art and decides to go there. After thinking for a moment she recalls that the Museum is on the 53rd Street. So, she walks there and enters the museum. The authors establish that Inge believes that the museum is on the 53rd Street and that she believes this even before she consulted her memory. The belief was somewhere in her memory and just waited to be accessed. The second person, Otto, suffers from Alzheimer’s disease. Like many similar patients he relies on information from the environment and so he always carries a notebook around with him. When he needs some old information, he just looks into his notebook, which for him plays a role of a memory. So, when he hears about the exhibition at the Museum of Modern Art and decides to go there, he just consults his notebook. It says that the Museum of Modern Art is on the 53rd Street. So, he walks there and enters the museum. Clark and Chalmers think that both cases are in relevant aspects analogues because the notebook plays for Otto the same role that memory plays for Inge. What counts is that information in the notebook functions just like information constituting an ordinary non-occurrent belief and it does not matter if this information lies beyond the skin. “The moral is that when it comes to belief, there is nothing sacred about skull and skin. What makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body.” [1; p.35].

It seems implausible that each notebook or perhaps even the whole Internet would count as part of my memory. But is it possible to prevent such excess? Clark and Chalmers pose a set of additional criteria to be met by non-biological candidates for inclusion into an individual’s cognitive system. They are summarized by Clark [13; p.46] in the following way:

- that the resource be reliably available and typically invoked (Otto always carries the notebook and will not answer that he “doesn’t know” until after he has consulted it),
- that any information thus retrieved be more or less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory,
- that information contained in the resource should be easily accessible as and when required.

In The Extended Mind article there is an additional fourth criterion that the information in the notebook has been consciously endorsed at some point in the past [1; p.38], but the authors subsequently drop it. They take these three requirements, called *glue and trust* criteria, as sufficient to rule out implausible candidates, as for example my shopping list.

Clark and Chalmers give another thought experiment which is more in line with HEC and involves three ways of playing the computer game Tetris [1; pp.27-28]. In Tetris, the player rotates falling blocks to form complete horizontal rows which are then eliminated. Imagine three cases:

**Case 1:** A person is sitting in front of a computer screen and must mentally rotate a block to align it with the sockets.

**Case 2:** A person is sitting in front of a computer screen and can choose either to mentally rotate a block as before or to physically rotate the image on screen by pressing a rotate button.
Case 3: A person is sitting in front of a computer screen and can choose to perform the rotation either by old-fashioned mental rotation or by using the neural implant that quickly performs the neural operation.

Clark and Chalmers suggest that all three cases are similar. First, Case 3 with the neural implant is just as much a cognitive process as Case 1. It seems there is no reason why an implant cannot count as cognitive just because it is artificial. And second, Case 2 is just as much a cognitive process as Case 3. One can imagine that Case 2 displays the same sort of computational structure as Case 3, although it is distributed across agent and computer instead of internalized within the agent. One cannot object that Case 2 is cognitive simply by pointing to the skin/skull boundary, since the legitimacy of that boundary is precisely what is at issue. So, if the rotation in Case 3 is cognitive, so is in Case 2.

This thought experiment suggests a kind of ‘offloading’ into the external environment. It can be an old fashioned pen and paper. For example, when we need to multiply high numbers, let say 455 and 678, we will use pen and paper and apply an algorithm we have learned in school. Or, we use computer technologies as for example brain-computer interfaces in order to access external databases, or to offload computationally-intensive processing.

They use this thought experiment as a springboard to offer a parity principle: “If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process.” [1; p.29]. Important are deep computational commonalities and not the way functions are materially implemented.

Nevertheless, it seems that the parity principle and the glue and trust criteria are much too liberal and insufficient to establish external elements as parts of one’s mind or cognitive system. As Palermos suggests, “if any external element that both satisfies the glue and trust criteria and causally affects one’s cognitive processes is to count as part of one’s cognitive system, we are going to be led to a ‘cognitive bloat’ ” [12; p.28].

COUPLING-CONSTITUTION FALLACY AND THE HYPOTHESIS OF EMBEDDED COGNITION

Adams and Aizawa criticize the argumentation that leads to HEC by pointing to the “coupling-constitution fallacy” [14-17], often also called causal-constitution fallacy [12]. The fallacy is committed when from the fact that some object or process is coupled in some way to cognitive agent, one slides to the conclusion that the object or process constitutes part of the agent’s cognitive apparatus or cognitive processing. They support the claim that coupling relations are distinct from constitutive relations by the following example: “The neurons leading into a neuromuscular junction are coupled to the muscles they innervate, but the neurons are not a part of the muscles they innervate. The release of neurotransmitters at the neuromuscular junction is coupled to the process of muscular contraction, but the process of releasing neurotransmitters at the neuromuscular junction is not part of the process of muscular contraction.” [15; p.68]. Or, giving a more general formulation, “we cannot assume that casually coupling a process X to a cognitive process Y is sufficient to make X a cognitive process” [14; p.93].

One possible way out is to weaken the thesis. Instead of arguing for the constitutive contribution of the external elements to one’s cognitive system, one should claim that cognition is many times merely dependent on external elements [12, 17]. This less radical hypothesis is called the hypothesis of embedded cognition (HEMC) and is defined by Rupert as “cognitive processes depend very heavily, in hitherto unexpected ways, on organismically
external props and devices and on the structure of the external environment in which cognition takes place” [17; p.393]. Although this hypothesis is close to HEC because it acknowledges that cognition is dependent on the external factors and the environment, it does not take these external factors as cognition’s proper (constitutive) parts. Cognition is restricted within the organism (brain or body, dependent on further theory) and cognitive mechanisms are internal, but explanations of the cognitive processes involve both internal and external factors. “According to HEMC, we can properly understand the traditional subject’s cognitive processes only by taking into account how the agent exploits the surrounding environment to carry out her cognitive work” [17; p.395]. In contrast, HEC implies that many times we should set aside our traditional subject because “the unit of analysis should be the organism and certain aspects of its environment treated together, as a single, unified system” [17; p.395].

So, HEMC is more conservative in retaining our common sense intuitions and it seems that it deals better with avoiding causal-constitution fallacy. HEMC maintains that cognitive processes causally depend on external tools and feature of the environment, while HEC maintains that cognitive processes constitutively depend on external tools and features of the environment [16; p.591].

Adams and Aizawa [14, 15] diagnose that Clark and Chalmers commit the fallacy because they do not specify what makes a process a cognitive process rather than a non cognitive process. They argue that if one takes any sort of information processing as cognition, it is likely that cognitive processing will be crossing the brain, body and environment. Because information is propagated through media, hard discs, televisions and telephones would be implausibly considered as cognition. They admit that processing information is plausible construed as a necessary condition on cognition, but reject it as sufficient for cognition. Proponents of the HEC thus need a theory of the “mark of the cognitive”. But this is not an easy task and Adams and Aizawa themselves admit that there is no well-established theory of exactly what constitutes the cognitive. Based on what they see as the common praxis in cognitive psychology they provide two clues. First, cognition involves non-derived representations, representations that mean what they do independently of other representational or intentional capacities. Second, cognitive is to be individuated by specific kinds of information processing mechanism that is located in the brain [14; p.31].

It seems to me that claiming that cognitive processes are implemented in the brain, and, at the same time, only those processes that are implemented in the brain count as cognitive, is begging the question. This way the chances for extended cognition are ruled out without further consideration. But on the other hand, Adams and Aizawa rightly point out that the advocates of HEC need to specify what they take as the mark of the cognition. If they do not provide an alternative proposal I feel that HEMC is in a better position. I will show in the continuation, there is a promising suggestion how to escape cognitive bloat and avoid coupling-constitution fallacy. It is based on the principle of complementarity and suggests specific connectionist models and dynamical system theory approach.

**TWO WAYS OF THINKING ABOUT HEC**

I think a lot of uneasiness about extended cognition comes from the fact that the discussions revolve mostly around the principle of parity [1] – the idea, that cognitive processes extend into the environment when some relevant parts of the world function the same way as the cognitive processes in the head. John Sutton has called discussions related primarily to the principle of parity, as the first wave thinking about HEC [18; p.193]. The principle of parity
stresses functional isomorphism between inner and outer processes and states. His argues that if *exograms* act as *engrams*, then for explanatory purposes they can be treated as engrams and the difference in their location is entirely superficial. “Thus breaking down classical and individualist distinctions between brain, body, and world, we see that the object can be (part of) the subject, and that, as we’ve noted, things can have a cognitive life” [18; p.193].

The first wave discussions based on the principle of parity assume functionalism and multiple realizability. Thus, when the critics are concerned with the distinction between inner and outer, they are not really interested in the differences in material realization. Adams and Aizawa [14, 15] point out that *intracranial* and *transcranial* processes are different with respect to the form of representations and their dynamics.

But the HEC is not supported only by the principle of parity. Clark has already in his book *Being There: Putting Brain, Body, and World Together Again* [19] explored how to include outer elements in order to make the whole system more efficient. The idea leads to the *complementary principle* and according to Sutton [18] originates the second wave of discussion. According to this principle “external states and processes need not mimic or replicate the formats, dynamics, or functions of inner states and processes. Rather, different components of the overall (enduring or temporary) system can play quite different roles and have different properties while coupling in collective and complementary contributions to flexible thinking and acting. So ‘exograms’ can be radically unlike engrams even while co-opted for the same purposes, and these differences will often be the focus of complementarity-oriented explanations” [18; p.194].

These second wave discussions can avoid some intuitively implausible consequences. For example, one does not need to accept the claim that cognitive states and processes are attributed to the external elements of the environment which can exist independently of humans (e.g. notebook). Also, artifacts do not work necessarily only as substitute for the brain via employing the same processes. The idea of extended cognition is thus not based on the principle of parity, but considers that body, artifacts and other external structures together with the brain form cognitive system that is enabled to perform different cognitive tasks as remembering, perceiving, language communication, learning and reasoning.

**PRINCIPLE OF COMPLEMENTARITY, CONNECTIONIST MODELS AND FEEDBACK LOOPS**

The first examples of cognitive models based on the principle of complementarity were specific connectionist models. Imagine a network where one can not store isolated atomic representations that can be further combined according to the rules. The advocates of cognitivism and the classical symbolic paradigm argued that because such models do not employ symbolic system and thus the representations lack the compositional structure (i.e. combinatorial syntax and semantic as employed in the language of thought), they can not explain some obvious features of thought as is systematicity. In short, thought is systematic when someone who can produce and understand the sentence “Mary loves John.” is also able to produce and understand the sentence “John loves Mary.” For cognitivist the ability is explained by the internal structure of the sentence [20]. I think critics are right when they point out that it is not possible to model systematic behavior by simple networks. But simple models do not exhaust all possibilities and scientists eventually design more complicated models that were able to learn such task. The subsequent “systematicity debate” closely resembles the current debate about extended cognition. Namely, internalists argue that such networks, although showing the required behavior, do not fulfil the task because the
Extended cognition: feedback loops and coupled systems

Explanation is not based on the internal structure [20, 21]. In contrast, connectionists stress that the network is able to learn to complete the task only if taken together with the external symbol system [22]. Solving the explanatory task for this approach thus consists in the division of labor between external symbols with combinatorial syntax and semantics and a system that is sensitive to them [23].

Although proposed connectionist models were relatively simple they showed crucial characteristics of a coupled system. Palermos argues for the postulation of a coupled system with two distinct arguments. “First, the properties that arise out of the interaction of coupled systems cannot be attributed to any of the contributing systems alone, but to the coupled system as a whole. … Second, in cases of ongoing feedback loops between coupled systems, there is a dense non-linear causal interdependence that disallows us to decompose systems in terms of distinct inputs and outputs from the one to the other” [12; p. 33].

The most important feature of cognitive system being genuinely extended is thus continuous reciprocal causation (CRC). This feature was already mentioned by Clark [24], but it seems that for him it is only a sufficient condition on cognitive extension and so examples like shopping lists are not ruled out. I think that Palermos [12] requirement that CRC is a sufficient and necessary condition successfully blocks the cognitive bloat and causal-constituency fallacy. But at the same time it rises the bar high and it is not easy to satisfy it. We have to bear in mind that the system is individuated on the bases of the process one is interested in and would be intuitively called cognitive. Such system will be called extended, if the task will be accomplished on the basis of continuous mutual interactions between the agent and his artifact, and will be at most embedded if these kinds of interactions will not be present [12; p. 34]. It remains open how to look at the example of Otto’s notebook. The answer depends on the context – how we imagine Otto is using it. If he is continuously interacting with it like one would with one’s own memory, then CTC criterion is satisfied. The reason why this may still seem strange is in the timescale. Most probably the real examples require feedback loops on much quicker scale. In a way, analogous processes happen in the brain where there are different feedback loops between different neural components.

CONCLUSION

I have explored the ideas that mind and cognition are not bound to the inner processes, most notably to the brain. At the beginning the idea of externalism appeared in the area of semantics and then spread to mental content and further to cognition and mind. Clark and Chalmers went further by arguing for active externalism and the extended mind. I have discussed the most common objections to this position. I suggest that there are better criteria than those originally proposed by Clark and Chalmers, namely continuous reciprocal causation and ongoing feedback loops. We are facing a quick development of artificial cognitive tools that have great impact on our cognitive performance. For example, we all feel that Google has changed how we look for information [25]. But it will take a detailed analysis of each concrete example to determine what can count as a cognitive system. If it will not be bound by the brain or maybe more liberal, to the skin, it will count as extended cognition. I am more sceptic, as were in fact also Clark and Chalmers, about qualitative experiences and feelings. I think we are still struggling to give them a proper treatment in naturalistic approaches [26-28]. Nevertheless I think that the hypothesis of extended cognition opens up the need for new interdisciplinary collaborations between biological, humanistic, social and technical approaches. “Thus, in seeing cognition as extended one is not merely making a terminological decision; it makes a significant difference to the methodology of scientific investigation. In effect, explanatory methods that might once have been thought appropriate
only for the analysis of ‘inner’ processes are now being adapted for the study of the outer, and there is promise that our understanding of cognition will become richer for it.” [1; p.30]. One such example is the investigation of the role of language as a tool in extending cognition [29].

Let me conclude the article with some remarks on potential ethical considerations of extended cognition. It is obvious that for many social and legal purposes, it is convenient to simply identify the agent with the core biological ensemble. “We imprison the body and brain, not the laptop!” [30; p.114]. But, as Clark [30] continues, also individual bits of neural circuitry, for example hippocampus, are themselves as incapable of being guilty as the laptop. It is the whole pattern of behavior that has itself emerged from a whole social and biotechnological matrix [30; p.114]. We know that treating the mind and self as machinery which is identical to the machinery of conscious reason leads to the conclusion that free will is an illusion and consequently question that human capability of taking moral responsibility [31, 32]. I see the extended cognition as a much more plausible naturalistic approach to these issues. As Dennett has put it: “Our free will, like all our other mental powers, has to be smeared out over time, not measured at instants. Once you distribute the work done ... in both space and time in the brain, you have to distribute the moral agency around as well. You are not out of the loop; you are the loop” [33; p.242].

ACKNOWLEDGEMENT

I would like to thank my colleague Toma Strle for inspiring discussions.

The work has been supported by Slovenian Research Agency (research core funding No. P6-0252) and by the Croatian Science Foundation under the project number IP-2014-09-6963.

REFERENCES


