Synthese (2020) 197:2143–2170
https://doi.org/10.1007/s11229-018-1797-5

Extended mind, functionalism and personal identity

Miljana Milojevic[1]

Received: 10 July 2017 / Accepted: 27 April 2018 / Published online: 5 May 2018

Abstract In this paper, I address one recent objection to Andy Clark and David Chalmers's functionalist argument for the extended mind thesis (EM). This objection is posed by Kengo Miyazono, who claims that they unjustifiably identify the original cognitive subject with the hybrid one in order to reach their conclusion about the mind extension. His attack consists of three steps: (a) distinguishing hybrid from traditional cognitive subjects based on the systems reply originally directed at Searle's Chinese room argument; (b) pointing out that the conclusion of the functionalist argument for EM must be rephrased to state that there are hybrid, and not extended, systems with widely realized mental states; and (c) arguing that functionalist EM cannot justify the assumption about the identity of these two kinds of subjects without circularity. I argue that Miyazono's main argument is ill-founded but that it, nevertheless, points out one important issue, namely, that we need further justification of the identity assumption, without which EM loses much of its flavor. Thus, I am going to challenge Miyazono's argument, provide a reinterpretation of the argumentation in the EM debate, defend the possibility of wide and extended selves, and offer a justification of the identity assumption, which I find crucial not only for vindicating EM but also for differentiating EM from other similar theses, such as the thesis about group minds.

Keywords Extended mind · Systems reply · Personal identity · Functionalism

✉ Miljana Milojevic
miljana.milojevic@f.bg.ac.rs

[1] Department of Philosophy, Faculty of Philosophy, University of Belgrade, Čika Ljubina 18-20, Belgrade 11000, Serbia

1 Introduction

In their 1998 paper "The extended mind", Andy Clark and David Chalmers argue that sometimes mental states of cognitive subjects extend into the environment. They dubbed this claim the Extended Mind thesis (EM). In order to justify it, they offered the Parity Argument (PA) based on the premise—that (partly) external processes are cognitive if they function as processes which we would recognize as cognitive were they done in the head.

Various authors have tried to show that PA fails as an argument for EM in numerous ways. These counterarguments were mainly focused on the functionalist nature of its major premise. Critics argued that the version of EM which follows from the argument with such a premise would be absurd (Sprevak 2009), trivial (Walter 2010) or unjustified (Rupert 2004; Adams and Aizawa 2008), depending on how coarse-grained we take the relevant functional roles to be. On the other hand, in "Does functionalism entail extended mind?" Kengo Miyazono (2017) takes a new perspective on the problematic nature of PA, independent of considerations about the relevant functional roles. He claims that even if we treat PA as a valid guide to recognizing relevant states and processes as mental or cognitive, PA cannot secure the conclusion that cognitive subjects' minds extend.

His argument has three steps: (a) showing that hybrid systems[1] are distinguishable from original cognitive subjects; (b) pointing out that PA's conclusion must be rephrased to state that it is rather the hybrid system and not the original subject who is rightfully attributed with relevant mental states; and (c) explaining why functionalist EM does not have the means to justify the identity assumption that the hybrid and the original system are one and the same. The first part of the argument is based on the famous systems reply (SR), originally applied to Searle's Chinese Room Argument. The second one is a consequence of the application of SR to EM; while the third is based on the examination of a possible justification of the identity assumption and ends with the conclusion that functionalist attempts to provide such justification are circular. If Miyazono's argument were sound, it would lead to a weakening of the original EM thesis. EM would not be a thesis about extension, or a thesis about *us*, people, as wide cognitive systems,[2] but a less interesting thesis about a possibility of hybrid cognitive systems.

I believe that Miyazono's main insight is correct, namely, that we need further justification of the identity assumption. I also gladly concede that PA is insufficient to fully vindicate EM and that additional assumptions are needed to establish its plausibility. Nevertheless, I do not agree that "extended mental states are metaphysically impossi-

[1] Hybrid systems are wide cognitive systems comprised of biological organisms (or at least their relevant parts) and pieces of the organism's environment that jointly form a supervenience base for relevant mental states or cognitive processes. I use the expression "hybrid system" instead of "extended system", because it is neutral with respect to the claim that hybrid systems are extensions of traditional cognitive subjects, which is going to be challenged.

[2] The thought that EM is about "normal human agents" is continuously present in works of Andy Clark, he writes: "*I* am slowly becoming more and more a cyborg. So are *you*" (Clark 2003, p. 3, emphases added), "Otto *himself* is best regarded as an extended system" (Clark and Chalmers 1998, p. 18, emphasis added), etc.

ble according to functionalist theories of mind." (2017, p. 3540) I want to show that this radical conclusion is a consequence of a narrow understanding of functionalism and of the original intentions of the authors who introduced PA. I will start with the exposition of PA, followed by Miyazono's objection. After evaluating the strength and importance of his objection I will offer my criticism based on the reinterpretation of the functionalist argument for EM. This reinterpretation will include one possible functional justification of the claim that sometimes hybrid cognitive systems are identical with original cognitive subjects and thus present their extension. This will strengthen the original argument for EM and help differentiate it from other forms of vehicle externalism. Also, I will address two arguments against the possibility of extended selves (Olson 2011; Wilson and Lenart 2014), which can potentially jeopardize the view I am defending, and provide one possible account of wide selves.

2 The parity argument, functionalism, and the extended mind

The Parity Argument for the Extended Mind is based—as already noted—on a functionalist premise called the Parity Principle (PP) which states:

> 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. (Clark & Chalmers 1998, p. 8)

The principle was meant as a claim against neural chauvinism. To function as a premise in an argument for the existence of such extended states, it had to be supplemented with claims which affirm its antecedent, in other words, with appropriate examples of processes which "function as" unequivocally recognized cognitive processes. Clark and Chalmers introduced the scenario with Otto and Inga (to be dealt with shortly) to supply such an affirmation. Nevertheless, this scenario had a second, different role to play. Namely, although PP is clearly committed to some form of functionalism by emphasizing that some kind of functional isomorphism plays a crucial role in identifying processes as cognitive, it does not specify the level of grain or character of relevant functional roles, and in this respect it is rather vague. Thus, the character of appropriate functional roles was left to be illustrated by way of examples.

Here is the story. Otto and Inga differ in one important respect: Otto suffers from Alzheimer's disease, while Inga does not have any substantial brain plaques and tangles, and her biological memory is unimpaired. Furthermore, because of his affliction, Otto uses a notebook to store information which Inga, a healthy cognitive subject, stores in her biological memory. Otto and Inga are both contemporary art aficionados. One day the two of them independently learn that an interesting exhibition is taking place at the Museum of Modern Art (MoMA). After Otto learns about the exhibition he consults his notebook, which contains a written address of MoMA, while Inga consults her biological memory in order to retrieve the information about the museum's location. Upon retrieving the information, they both make decisions about the directions they should take. It is claimed that the information about the location of MoMA in Otto's notebook plays the same relevant causal roles as Inga's dispositional belief

stored in her biological memory. In other words, it is claimed that Otto's exogram (Donald 1991) functions in the same manner as Inga's engram.[3] If we assent to PP and to the claim that Otto and Inga have relevant functionally isomorphic states, we can infer that there are actual extended beliefs.

PA for Otto's extended belief is further supported by the additional specification of conditions which the relevant information has to satisfy. Clark (2010) later called these additional conditions "glue and trust". Their role was to secure further integration of the notebook into the hybrid Otto-notebook cognitive system and to safeguard against overextension. According to these conditions the notebook should be readily available to Otto, information Otto retrieves from it should be more-or-less automatically endorsed, the information he retrieves should be easily accessible and previously consciously endorsed.[4]

PA gained a lot of attention and the subsequent debate about its plausibility was mainly focused on the kind of functional roles needed to establish the relevant functional isomorphism between extended and internal states, although it was not reduced to this kind of criticism. I will use Gertler's (2007) refined reconstruction of the original argument to show which premises were most commonly under attack, and to illustrate the point of divergence of Miyazono's criticism from past objections.

(1) "What makes some information count as a [standing] belief is the role it plays" (1998, p. 14).
(2) "The information in the notebook functions just like [that is, it plays the same role as] the information constituting an ordinary non-occurrent belief". (1998, p. 13)
(3) The information in Otto's notebook counts as standing beliefs. (from (1) and (2))
(4) Otto's standing beliefs are part of his mind.
(5) The information in Otto's notebook is part of Otto's mind. (from (3) and (4))
(6) Otto's notebook belongs to the world external to Otto's skin, i.e., the 'external' world.
(7) The mind extends into the world. (from (5) and (6)) (Gertler 2007, p. 193)

[3] While engrams are hypothesized physical traces of external stimuli in the brain, which constitute the physical basis of our memories, exograms are external symbolic memory devices which enhance our biological memory. The term "exogram" was introduced by neuropsychologist Merlin Donald.

[4] In the same article Clark and Chalmers provide an argument for the extension of differently identified states and processes, namely, cognitive processes. They argue that a task of assessing the fit of a tetremino block in the game of Tetris can be typically done with a help of two different processes—mental rotation of the block, or manipulation of the buttons on a Tetris console or a computer. The first one is normally perceived as a cognitive process while the second one is not. By pointing to the possibility of a possession of a neural implant which functionally performs in a similar fashion as the physical manipulation of Tetris buttons, and assuming that we would be inclined to call this neural implant a part of our cognitive machinery, they conclude that by the Parity Principle we should regard the manipulation of the buttons as a part of the cognitive process. The only difference between the button manipulation and the neural implant performance is in a manner of their realization, one is external and the other internal, while their functions are identical. In sum, the Parity Principle gives grounds for arguing for the extension of both mental states and cognitive processes yielding two conceptually close hypotheses: EM and EC (Extended Cognition).

Only premise (6) has escaped dispute. Premise (1) was most notably criticized by Shapiro (2004), who claims that functionalism as a theory of mind is inadequate, that the mind is deeply incarnated and that is inseparable from the body. Premise (2) has received the most attention. Rupert (2004) and Adams and Aizawa (2001, 2008) insisted that functional/causal roles needed to identify a state as a belief are not played by the extended state because they are more fine-grained than those singled out by Clark and Chalmers (e.g. biological memory is subject to priming, recency effects, has limited storage, etc.). Sprevak (2009), on the other hand, tried to show that if there are functional roles realized by both extended and internal processes then they are so coarse-grained that they would be also satisfied by absurdly overextended processes, e.g. processes of a system formed by me and a program on my computer. Gertler (2007) herself questions premise (4), by insisting that only occurrent beliefs constitute a mind. Her argument is similar to Sprevak's as it was designed to show that PP leads to overextension (this time to overextension of agency). Nevertheless, her conclusion is not that functionalism should be re-evaluated, but rather that we should reject the assumption that standing beliefs constitute a mind as a less costly alternative.

The novelty of Miyazono's approach is his focus on subconclusion (5), which he claims is unwarranted. He argues that (5) does not follow from (3) and (4), because there is, in his view, a further alternative—relevant beliefs can be Otto-notebook system's beliefs, and not Otto's (3'). We can, thus, successfully argue for Otto's mind extension only if we eliminate this alternative, or if we show that the Otto-notebook system is identical with Otto (OAS). Miyazono believes that elimination of (3') is not an option. He argues for (3') by adding the systems reply (SR) to subconclusion (3). Thus, after the introduction of SR, subargument (3–5) has to be reformulated in the following manner:

(3) The information in Otto's notebook counts as standing beliefs. (from (1) and (2))
(3') Standing beliefs realized in the notebook are the Otto-notebook system's beliefs. (from SR and (3))
(OAS) The Otto-notebook system is Otto.
(3") Standing beliefs realized in the notebook are Otto's beliefs. (from (3') and OAS)
(4) Otto's standing beliefs are part of his mind.
(5) The information in Otto's notebook is part of Otto's mind. (from (3") and (4))

The validity of the argument is now restored, but according to Miyazono the argument is not sound. Namely, he argues against OAS, which leads to rejecting the claim that standing beliefs realized in the notebook are Otto's beliefs (3"), and consequently to rejecting both (5) and (7), thus rebutting PA. According to this objection, PA can be, at best, rephrased to claim that there are hybrid, and not extended, systems with standing beliefs. In Sect. 3 I will address Miyazono's justification of (3'), or more precisely justification of ONSB which is a special case of (3'), and in Sect. 4 his rejection of OAS.

3 The systems reply to the extended mind

Miyazono claims that PA can become a sound argument only if its (sub)conclusion is rephrased to state that the belief about the location of MoMA realized in the notebook is not Otto's belief but the Otto-notebook system's belief (ONSB)[5]. Miyazono bases his argument for this claim on SR which emerged in the debate about Searle's Chinese Room Argument (1980).

Searle's original argument was designed to show that manipulating symbols according to formal rules is insufficient for understanding. The argument is based on a thought experiment. Imagine a person with no prior knowledge of Chinese, locked in a room with a batch of Chinese symbols, a set of rules written in his native tongue, and a slot through which this person could receive sentences written in Chinese, and return answers written with a help of the Chinese symbols stored in the room. Searle built his skeptical conclusion on the intuition that we can imagine that this person manages to give perfect answers in a language she does not understand only by the manipulation of formal symbols according to the rules. The scenario was designed to show that the computationalist assumption about the mental is false. The mental is not reducible to computations done over symbolic representations, because, as the scenario shows, there is something left out from the computationalist story. Intentionality and understanding require semantics, which is not given in formal descriptions of computational algorithms.

Several commentators have agreed with Searle that the person in the room would lack an understanding of Chinese, but maintained that he is wrong to conclude that there is necessarily no understanding in the case described. The objection points out that the person in the room is not the appropriate subject to whom we could ascribe relevant understanding according to computationalism. The person in the thought experiment is only a central processing unit (CPU) of a larger system. Every computer needs, in addition to a CPU, at least a memory and a set of instructions. To evaluate the mental and cognitive status in this case we have to take into account a broader system comprised of the person in the room, the written rules, scratch paper used for calculations, and "data banks" of sets of Chinese symbols (Searle 1980). It is only this wider system that can be attributed with an understanding of the questions that are fed into it, and of the answers it provides in Chinese. This response came to prominence under the name "Systems Reply".[6]

[5] ONSB is a special case of (3'). Gertler generalized relevant claims about the particular belief about MoMA to all standing beliefs realized in the notebook (in accordance with (1) and (2)), thus the difference between our interpretation of PA and its premise (3'), and Miyazono's interpretation and ONSB. Because this generalization is justified by (1) and (2), Miyazono's conclusions about the application of SR to EM can be accordingly generalized from ONSB to (3').

[6] Searle (1980) associates this kind of response to Berkeley. The original response identified the relevant subject with the described system, and it was challenged by Searle's objection that even if the person in the room internalizes the database and the instructions, she would still be unable to understand Chinese. Later versions of SR, sometimes called Virtual Mind Reply (see Cole 1991), do not identify the relevant subject with the system, but rather claim that this subject is realized by the system. One system can realize two or more subjects or persons, thus this version of SR avoids the internalization objection. The person in the room after memorizing the instructions, and the database, would be able to simulate a separate person who understands Chinese. Both of these versions are compatible with arguments that follow.

Miyazono argues that if we adopt the insight of SR, that it is only the hybrid system, as described, which is capable of relevant understanding, we should make a reference to system's components in attributing relevant mental states or cognitive processes responsible for the realization of these states. In other words, if we accept PP (1), and take the Otto-Inga scenario as a good example of extended processes (2), we can infer:

> **ONSB1**: The belief that the museum is on 53rd Street is physically realized in the notebook.
>
> By applying SR to Otto's case we get:
>
> **ONSB2**: It is the Otto-notebook system, rather than Otto, that believes that the museum is on 53rd Street.
>
> And, finally, the conjunction of these two claims is:
>
> **ONSB**: The Otto-notebook system's belief that the museum is on 53rd Street is physically realized in the notebook.[7] (Miyazono 2017, p. 3525)

But even if we concede that the relevant belief should be attributed to the Otto-notebook system why should we consider this system as different from Otto? To further his point, and to establish a difference between the hybrid and the traditional subject, Miyazono uses Sprevak's controversial example (for criticism see Drayson 2010; Wheeler 2010), which was originally introduced to show that EM leads to absurd conclusions.

Sprevak asks us to imagine the following:

> There is a Martian whose memory, instead of being stored in patterns of neural activity, is stored internally as a series of ink-marks. […] As well as acquiring beliefs via its senses, the Martian is born with innate beliefs that it has not examined yet—a library of data that is hard-wired into the organism by developmental processes. (2009, p. 508; Miyazono 2017, p. 3528)

Intuitively, Sprevak claims, the information in the Martian's storage counts as his dispositional beliefs. We can further imagine that the Martian, in his storage, has a belief that Alfa Centauri is bigger than the Sun. On the other hand, if we accept PP and find an example of a state that plays the same functional roles as the Martian's belief about the comparative size of the two stars, this state should also be considered as a dispositional belief. Sprevak continues with such an example. Back on Earth, a person named Mark owns a book that records the same information. Mark has not examined this information yet. But according to PP, in Sprevak's opinion, we should regard the information in Mark's book as Mark's belief, which is absurd.

Miyazono does not agree with Sprevak's conclusion that the version of EM that follows from PA is absurd, and reinterprets Sprevak's scenario in accordance with SR. If we apply SR to the Mark-book example it becomes an illustration of a case where we can grant a mental status to the information in the book, but at the same time we should not attribute this mental state to Mark, as Sprevak does, but to the Mark-book system. This reinterpretation demonstrates that even if it is absurd to attribute the

[7] ONSB1 is a special case of (3) "The information in Otto's notebook counts as standing beliefs"; ONSB2 is the consequence of the application of SR to Otto's case; and ONSB is a special case of (3').

information stored in the book to Mark as his dispositional belief, it would not be absurd to attribute it to the Mark-book system.

The intention behind the application of SR to the Mark-book scenario is to show that hybrid systems are distinguishable from their originators and that they are not to be confused. Their potential identity has to be independently justified. So the introduction of this example was meant as a basis for establishing a clear difference between biological and hybrid systems based on differential attributability of mental states, which plays a crucial role in Miyazono's rebuttal of PA. It was not sufficient to point out that we have to make a reference to component parts of the system as requested by SR, it was necessary to show that this system is also different from the traditional cognitive subject, as in the Mark-book case. The same should apply, by analogy, to Otto and the Otto-notebook system. In other words, they are to be considered different if not proven otherwise.

4 Implications of the systems reply objection

Before I criticize Miyazono's argument let us first evaluate the importance of his objection. As we saw, PA makes a claim about when a state or a process should be regarded as cognitive, and tacitly assumes that the state in question is Otto's state. This assumption—about Otto being the subject of the relevant state—appears in the description of Otto-Inga scenario, but it is not independently supported: "Otto believed the museum was on 53rd Street even before consulting his notebook" (Clark and Chalmers 1998, p. 13). Let us remember how the argument goes:

> (1) "What makes some information count as a [standing] belief is the role it plays" (1998, p. 14).
> (2) "The information in the notebook functions just like [that is, it plays the same role as] the information constituting an ordinary non-occurrent belief". (1998, p. 13)
> (3) The information in Otto's notebook counts as standing beliefs. (from (1) and (2))
> (4) Otto's standing beliefs are part of his mind.
> (5) The information in Otto's notebook is part of Otto's mind. (from (3) and (4)) (Gertler 2007, p. 193)

Miyazono's objection is directed at the inference to (5) from (3) and (4). It was argued that in order to save the validity of the argument with conclusion (5) we have to justify hidden premise (3"): Standing beliefs realized in the notebook are Otto's beliefs.

The original omission of the explicit statement about the attribution of the dispositional belief to Otto seemed as a natural thing to do before the introduction of SR. There were only two possibilities: either the belief state was Otto's or the notebook's. Among these alternatives we easily picked out Otto, because notebooks do not have mental states or the ability to process them. Nevertheless, after the introduction of the third candidate—the Otto-notebook system—we have to account for why the relevant state is rather Otto's and not the system's belief. But first, we should investigate whether the introduction of this third alternative really jeopardizes EM, or just creates

a verbal dispute. Defenders of EM could accept ONSB (or (3')), say that they were not interested in questions about personal identity and that the crux of the thesis was just that there are states and processes that are in part externally realized, without regard to the continuity of the two systems. The EM thesis would be equally justified even if it was restated in a way that was suggested. However, it soon becomes clear that this is not just a verbal dispute, and that accepting that Otto is not the relevant subject in the story is incoherent with other assumptions of EM.

Namely, the idea that the identity of the subject stays unaffected in mental and cognitive extension, in a sense that the subject retains her identity, can be seen as one of the crucial assumptions of the EM hypothesis. If the widely realized states belonged to a different, separate, or a newly created entity (if hybrid Otto were distinct from purely biologically constituted Otto) we would not be able to say that the boundaries of *human* cognition have changed. Wide mental states would be states of newly constituted entities and not the states of *extended* minds. So, if defenders of EM "embrace SR"[8] their thesis will become much less attractive. It would not be a thesis about our cognitive and mental extension or about the change of our cognitive capacities by the use of artifacts. It would become a much less interesting and radical thesis, more similar to the hypothesis of strong AI, stating that there could be systems partly organic and inorganic whose states can play relevant causal roles.

Thus, arguing for (3') is not equivalent to arguing for EM. We have to claim OAS, too, and consequently (3"). On the other hand, if we simply add OAS as a missing premise, we face a threat of circularity. The argument for EM would rely on this assumption and at the same time the conclusion of the argument would be used as a support to argue for OAS or the extension of the self. This is illustrated with Clark and Chalmers's claim: "Does the extended mind imply an extended self? It seems so." (Clark and Chalmers 1998, p. 18; Miyazono 2017, p. 3535) In other words, the claim about the extension of relevant states would be based on OAS, and OAS would be defended by the proven extension of relevant states.

So it seems that we are facing a dilemma. Either we abandon OAS and make EM an uninteresting thesis, or endorse it and make the argument for EM circular. Of course, one might also offer an independent justification of OAS, but Miyazono believes that there is no functionalist justification of it. His investigation of a possible justification of OAS, though, is less than extensive. He considers the "glue and trust" conditions (Clark 2010) as possible candidates and concludes that they cannot do the trick.[9] He dispenses with them by arguing that they "are, however, very controversial. For instance, Sprevak (2009) points out that the cognitive resources involved in acts of outstanding human creativity are not reliably available or typically invoked" (2017,

[8] Miyazono talks about "embracing SR" as one of the possible answers to his objection. He examines three such possible answers in his paper: defending the identity of Otto and the Otto-notebook system (or defending OAS), claiming that there are no selves, and embracing SR. For the purposes of this paper, I am going to address only the first and the last of these possible responses.

[9] Namely, if the "glue and trust" conditions could make a difference between the Otto-notebook system and the Mark-book system, perhaps they could be considered as sufficient for justifying OAS. Recall that arguing in favor of the difference between hybrid and biological systems was based on Sprevak's example. So, if the Otto example is not analogous to it, then it might turn out that he is not different from the Otto-notebook system.

p. 3536). In the end he adds that this is not a conclusive argument that there is no independent justification of OAS, but "that, although OAS is a theoretical possibility, finding an independent argument for it is not very easy." As we can see, even Miyazono does not believe that his reasoning is very persuasive, but still insists that he has

> great difficulty in conceiving even a single case where, according to some sort of functionalism, OEB[10] (or a corresponding claim) is true. It is quite difficult to see how we can say that it is Otto, rather than the hybrid system, who believes that the museum is on 53rd Street. (2017, p. 3540)

In other words, the burden of proof is left on the EM side. So let us carry that burden for a little while, and try to provide an independent justification of OAS.

5 Vindication of the extended mind

5.1 The structure of the argument for EM

To answer Miyazono's criticism I first want to establish the exact position of PA and "glue and trust" conditions in the structure of the argument for EM. I believe that the vagueness of PP's "function as" and the interpretation of PA as a complete and definite argument for EM are the main culprits for numerous attacks that are launched at it, even almost twenty years after its formulation. I will reconceptualize the argument for EM, which will enable us to (a) isolate its problematic nature, (b) properly position Miyazono's objection, and (c) give one possible answer to this objection.

I distinguish three parts, or three steps, in the functionalist argument for EM:

(1) Showing that there are hybrid integrated systems.
(2) Providing an argument that sometimes these systems have mental properties.
(3) Justifying the claim that these systems are, at least sometimes, identical to subjects which are extended by processes of hybridization.[11]

I will now try to justify this tripartite division. I am going to provide reasons for introducing these steps and investigate the hypothesis that they can be accommodated in functionalist argumentation for EM. If this hypothesis is correct, OAS will be independently justified and SR objection rejected.

5.2 Hybrid integrated systems

Let us first examine step 1). I agree with Rupert (2009) that functional criteria alone are not sufficient for calling something cognitive, or at least in the form they are

[10] "OEB is the conjunction of two claims: **OEB1**: The belief that the museum is on 53rd Street is physically realized in the notebook. **OEB2**: It is Otto who believes that the museum is on 53rd Street." (Miyazono 2017, p. 3524) .

[11] I claim that step 1) and 2) are parts of the standard reasoning of the proponents of functionalist EM, although this is not always recognized. I add step 3), inspired by SR to EM.

usually presented.[12] In order to call a process cognitive or a certain state mental it has to be a part of an integrated system with appropriate properties. Typically, cognitive explanations are explanations of mechanisms of intelligent behavior and there has to be something that behaves intelligently: an agent or a subject capable of producing such behavior. This assumption of an underlying system is implicitly endorsed in most, if not all, theories of mind. Descartes postulated *res cogitans*, materialists focused on biological organisms and brains, computationalists on computational architecture, etc. While this assumption about an integrated entity or a structure capable of intelligent behavior comes in a variety of forms and with different metaphysical commitments, I take it in its most general form without committing to any robust metaphysical notion of the self or what Dennett (1991) calls the "central meaner". I am only committed to the existence of a system which enables causal interactions between mental states and cognitive processes in a way that can yield intelligent behavior and agency. Though I do not agree with Rupert that such systems coincide with biological organisms I emphatically support the notion that "the persisting cognitive subject is a theoretical construct, a hypothesized set of relatively stable cognitive capacities, states, or abilities (or their underlying mechanisms)." (Rupert 2009, p. 51) Showing that relevant systems are appropriately integrated is a crucial step in arguing that they are capable of having mental states.

But is it not the case that functionalists, including Clark and Chalmers, skip this step and make a mistake in taking only rough functionalist criteria into account? Certainly not. If we take a look at traditional functionalist theories it becomes clear that they employ integration assumption too. In Lewis (1980) it did not have to be stated explicitly, because it was implied. The functional properties by which we identify mental states are, according to him, realized in members of a population or a suitable natural kind. Lewis considers humans and hypothetical Martians, thus starts with entities recognized or hypothesized as biological systems, and continues to determine what kind of properties make them suitable for the attribution of mental states. He is, also, adamant in his opinion that mental states cannot occur in isolation. They have to be connected with other mental states in order to produce suitable behavior of the system of which they are a part, and which enables the needed interaction of these states in the first place.[13]

Nevertheless, if we are investigating hybrids and their states, as it is done in the research of EM, we cannot rely on the systemic organization of biological organisms.

[12] Functional roles used for identification of mental states as they are often conceived in the philosophy of mind literature; simplified examples which single out typical causes and effects of a certain type of mental states with vague indicators of causal connections with other mental states that reflect their inferential relations. The use of such simplified examples in presenting functionalism is one of the main reasons for separating considerations about integration from those about common-sense functional roles, although, as it is going to be argued, integration assumption, at least in one of its forms, is implied by the functionalist theory of mind.

[13] In fact, both Putnam, in his defense of machine state functionalism (1960), and Lewis (1972) characterize mental states holistically—Lewis by asking for the Ramsification of the whole psychological theory, Putnam by specifying every mental state with respect to the total state of the system. Although holism introduced new problems for functionalism, e.g. too chauvinistic characterizations of mental states, coarse-grained functionalism, even if not asking that every mental state should be defined in a relation to every other, kept the assumption that mental states are mutually connected and that they jointly produce appropriate behavior.

We have to offer an independent account of their integration, and an answer to the question: What makes a part of the world a part of a cognitive *system*? In much of the literature on EM, great importance was given to this question, and this is why I consider it as an equal part of the argumentation for EM together with parity considerations.[14] The request for integration is spelled out in different forms through the notions of "non-trivial causal spread", "dynamical coupling", "distributed functional decomposition", "continuous reciprocal causation", "glue and trust conditions", etc. (Clark 2008), which reflect various aspects of integration and describe integration on different levels. Integrated cognitive systems are integrated physical systems, meaning that their parts are mutually coupled and exhibit dynamical and causal regularities in their interactions, but they also exhibit typical forms of cognitive integration in terms of the cognitive functions that different kinds of flows and transformations of information perform. Sometimes, these systems will also have mental states which will be interconnected in a systematic way reflecting specific psychological functions and regularities. The relation between different levels of integration is such that lower-level integration grounds and partially explains higher-level integration, while higher-level integration is a sign of lower-level integration.

The notion of "distributed functional decomposition" is especially useful for our current purpose of characterizing cognitive integrated systems. This concept emphasizes the explanatory role of functional higher-order descriptions of integration and captures relevant features based on which we judge if we are encountering a system or just a collection of things.

> Distributed functional decomposition is a way of understanding the capacities of supersized mechanisms (ones created by the interactions of biological brains with bodies and aspects of the local environment) in terms of the flow and transformation of energy, information, control, and where applicable, representations. The use of the term *functional* in distributed functional decomposition is meant to remind us that even in these larger systems, it is the *roles* played by various elements, and not the specific ways those elements are realized, that do the explanatory work. (Clark 2008, pp. 13–14)

Thus, it is a specific function that different parts of the world perform that makes them a part of the same system. Nevertheless, it is important to note that conditions of particular cognitive integration can vary, and integration can occur in a number of ways depending on the contingent circumstances of a particular cognitive task or a function. The integration will depend on a number of dimensions and they will include:

> the kind and intensity of information flow between agent and scaffold, the accessibility of the scaffold, the durability of the coupling between agent and scaffold, the amount of trust a user puts into the information the scaffold provides, the degree of transparency-in-use, the ease with which the information can be inter-

[14] It is redundant to say that I see conditions and arguments for integration as complementary to PA as a functionalist argument for EM, and not as rivalry as they are conceived by some authors (Menary 2007; Sutton 2010). This was already suggested by the tripartite division of the argument for EM.

> preted, amount of personalization, and the amount of cognitive transformation. (Heersmink 2017, pp. 433–4)

In Otto-notebook case we find that Otto and the notebook form a system based on the nature of causal connections between them, characteristic feedback loops that obtain in their interaction, and the ways the information in the notebook is created, revised and used in the system. The notebook is a constant in Otto's life and it has become "transparent equipment" to him (Heidegger 1927/1961; Clark 2008), it became an integral part of his cognitive functioning. Furthermore, in Otto's case integration of the notebook is specifically explained by the "glue and trust" conditions. The information in the notebook was previously consciously endorsed, it is readily available, it is more-or-less automatically endorsed after retrieval, and it is easily accessible. These conditions describe the systematic character of stored informational states of the system, and how we expect dispositional beliefs to function in an integrated cognitive system.

Without going into further detail we can conclude that there are minimal general integration conditions—a request for the existence of a coupled system or reciprocal causal connections—and more specific conditions specified in accordance with purposes it ought to serve and functions it performs.[15] This implies that it is not enough to point out that specific conditions of integration are not met in some cases in order to prove that they are not conditions of integration at all. For this reason, Miyazono's dispensing with the "glue and trust" conditions as criteria for integration in Otto's case based on the insight that "cognitive resources involved in acts of outstanding human creativity are not reliably available or typically invoked" (2017, pp. 3536–7) is unjustified. I also want to conclude this subsection with the claim that the integration assumption needs to be independently justified, although it may be argued that it is already a part of functionalism even if it is often neglected. The integration assumption is particularly needed in the current time of imperfect science. Namely, if we were to have a complete psychological theory of the mental, namely, functional specifications of every possible mental state and complete descriptions of their mutual causal connections, perhaps we would not need an integration assumption because only appropriate systems would have states that satisfy all of the specified causal roles.

I should also briefly note that there are no similar conditions of integration offered for the Mark-book system, and the description of that case does not give us reasons to believe that any such conditions could be at all specified. The Mark-book system is not a system at all; it is better specified as a collection of objects, than as an integrated entity. This kind of criticism is elaborated in Wheeler's (2010) response to Sprevak, who examines a different but analogous example that Sprevak uses to reach his conclusions about overextension, namely his description of Mayan calendar program stored on his computer and in the Martian's head. Wheeler argues that we should reject the cognitive status to both implementations of the Mayan calendar program because they

[15] For more detailed treatment of cognitive integration see Menary (2007). Menary distinguishes between four types of bodily-environmental cognitive integrations: bodily couplings, epistemic actions, self-correcting actions, and cognitive practices.

are not properly integrated in either of these cases. The fact that the program is stored in the creature's head does not qualify it as cognitive. The thesis against locational chauvinism works both ways, something is not non-cognitive just because it is not in the head, but also something is not cognitive just because it is in the head (Wheeler 2010, pp. 19–20). The same should apply to the example that refers to the information about the comparative size of Alfa Centauri and the Sun.[16]

5.3 Mental properties of hybrid integrated systems and a rebuttal of the SR to EM

I have now considered step 1), so I shall proceed with step 2)—providing an argument that sometimes hybrid integrated systems have mental properties. This step is crucial in arguing for EM, and it is commonly taken as the argument's only part. I am not going to consider it in detail, because I find it already sufficiently explained. This step explicitly employs PP which should guide us in the identification of mental states of integrated systems by referring to common sense functional roles, and I described it in Sect. 2 as PA.[17] It is important to note that neither step 1) nor step 2) is solely sufficient to argue for EM. Step 2) is insufficient because of the reasons I provided in the previous subsection (imperfect state of current psychology, interactive character of mental states, etc.). The omission of step 1)—disregarding the assumption about systemic properties of mental states—leads to many unfavorable interpretations of functionalism. Namely, if we take only functional roles of a single mental state, in isolation from many other mental states with which it interacts, as its sole identify-

[16] Bearers of this information do not play appropriate roles in the "flow and transformation of information", because they are not "intimately embedded in subtle and complex perceptual, memory and reasoning systems that have been evolved or developed in relation to each other, and that already meet whatever the criteria are for cognitive status". (Wheeler 2010, footnote 8) This lack of cognitive integration thwarts the possibility of mental integration, because the dispositional beliefs with relevant informational content will not be able to play expected (postulated by common sense psychology) causal roles with other mental states of appropriate contents. Informational states in Martian's head (and in Mark's book) are not perceptually nor inferentially produced, they are not affected by the change of other informational states in the system, etc. The only reason given to consider them as cognitive or mental was that they are stored in the Martian's head.

[17] If we read PP wide enough it could be considered as supporting both (1) and (2), more precisely, examples used for the affirmation of its antecedent can be seen as confirming both integration of the relevant system and the mental status of its parts. The Otto-Inga scenario would show that information in Otto's notebook "functions as" a dispositional belief because the notebook plays integrative functional roles in Otto's mental life, and because specific information in it plays specific common-sense functional or causal roles used for identifying it as a particular kind of belief. The vague indicator "function as" used in the formulation of PP would refer to both partial functional roles used for the identification of mental states, and integrative functions, in Otto's case the "glue and trust" conditions. This is why we are considering PP together with specific examples which support it (or as PA), as their role is to specify the kind of functionalism which is to be used in determining functional equivalence. Without further specifications PP is neutral with respect to the kind of functionalism that is to be used to establish mentioned equivalence, and it is compatible with both common sense functionalism and scientific or psycho-functionalism about the mind, as well as with functionalism about cognitive processes (PP combined with such functionalism leads to EC). Nevertheless, in defending EM, Clark and Chalmers commit themselves to common sense functionalism as the relevant kind of functionalism which can establish it.

ing criterion, we are going to end up with many unfortunate realizers of so defined mental states. Information in an encyclopedia or a smartphone could play such partial causal roles. The information from these storage devices could be retrieved in appropriate circumstances and cause appropriate behavior, but we would not be prepared to grant them mental status. In EM literature this problem is known as the problem of overextension or as the problem of "cognitive bloat" (Adams and Aizawa 2001). In my view, this problem is sidestepped if we account for other causal connections of these information states by referring to their position and role in a system. Namely, instead of explicitly describing all other states with which, for instance, information in Otto's notebook is causally connected (that it is produced by Otto's occurrent belief about the location of MoMA, his wish to remember this information, etc.), we can point out that all of the information in the notebook is produced in the appropriate manner—by suitable conscious states of a relevant subject—that it is used by the system without intermediating judgment about its trustworthiness, etc. Also, the systemic role of Otto's beliefs can be further supported by showing in which ways the notebook is coupled with Otto. On the other hand, step 1) is insufficient for EM because it only shows that an external part has a role in an overall hybrid system. There is still no guarantee that any of its states can be identified as common sense mental states. This is why we have to employ PA[18].

Before I consider the justification of (3), which seems to be the step that Miyazono attacks, let me re-evaluate his general argument against EM based on SR. The argument was based on a possible difference between traditional and hybrid cognitive systems. It was claimed that we should differentiate between Otto and the Otto-notebook system, and that the extended belief states should be attributed to the Otto-notebook system. This claim was based on the analogy with the difference between Mark and the Mark-book system that was established on differential attributability of beliefs—some beliefs, it was claimed, are attributable to the system but not to Mark alone. The attribution of mental states to both Mark and the Mark-book system was led solely by the conditions stated in step 2). We can conclude now that this argument is flawed. The difference between traditional and hybrid cognitive system was ill-founded because we cannot attribute belief states to the Mark-book (system) at all. These two things—Mark and the book—are not sufficiently integrated and do not satisfy conditions of step 1). The Mark-book system, thus, does not have any extended mental states. This means that we were not offered good reasons to believe that Otto is different from the Otto-notebook system in the first place. The introduced analogy does not hold. Two kinds of cases are not similar in relevant respects (hybrid Otto is integrated, whereas Mark and the book are not), and also the claimed property of the source domain—differential attributability—does not exist. So, given this, should we proceed further? I believe that we should. Although the example we were given to raise our doubt about the identity of the two cognitive systems was not suitable, there could be other examples where this difference really exists. I still believe that the question "What makes a biologi-

[18] To illustrate this claim it is sufficient to point out to the difference between EM and EC (see footnote 4), and that EC does not entail EM. Namely, there could be extended cognitive processes, which do not instantiate any mental states recognized by common sense functionalism.

cal being and a hybrid identical?" is a valid question. In other words, I believe that Miyazono has pointed out a weak spot in the argument for EM, that there is no explicit justification of the identity claim, which is needed for the full blown thesis about the extension. Unfortunately, he did not build a sound argument to justify his claims.

For a moment let us consider extreme cases of (fictional) hybridization, as portrayed in the movies *RoboCop* or *Transcendence*. Perhaps it is appropriate to ask whether Alex Murphy and Will Caster are still among us after they have been merged with robotic or digital resources. Are mental states of these hybrids still mental states of the original Alex Murphy or Will Caster? Those questions are questions about personal identity, and a justification of OAS should be reconceived as a justification of the retention of Otto's personal identity. OAS claims that there is an identity between Otto and the Otto-notebook system which certainly cannot be physical identity. It can be sensibly interpreted only as a claim about the numerical diachronic identity of a person. In other words, as the claim that the-biological-Otto-at-T_1 is identical to the-Otto-notebook system-at-T_2. If personal identity is sustained through the process of hybridization then mental states which are realized in both biological and artifactual parts of the system are attributable to the person that entered the process of hybridization and thus constitute a part of the extended mind. I believe that this question cannot be answered a priori and that the retention of the identity through hybridization is a contingent matter. This is why I still feel obliged to give positive reasons for considering Otto to be the Otto-notebook system (justifying OAS), and include step 3) in the argument for EM.

5.4 Identity assumption

Although there are different views on persistence conditions of personal identity, I will, for now, uncritically assume one widely accepted view, namely, the psychological continuity approach. This view was inspired by John Locke's (1689/1975) understanding of personhood and personal identity, and it is contrasted with biological views, which seek to identify persons with their biological bodies. In Locke's view a person is a "forensic term appropriating actions and their merit; and so belongs only to intelligent agents, capable of a law, and happiness" (Locke 1689/1975, Book 2, Chapter XXVII). He considers a thought experiment in which a prince and a cobbler switch their bodies, and argues that what matters for personal identity is not any part of the body (given that it is possible to transfer thoughts without transferring or altering any part of the body), nor the identity of the soul (given that souls exist as substrates of thoughts), but the continuity of thoughts and memories of past experiences. This major premise of the psychological continuity view makes it essentially functionalist.[19] In present times this view is motivated by the changing nature of our own bodies and possibilities of brain or mind transfers. According to Parfit:

[19] By calling this view "functionalist" I do not intend to claim that Locke's view is compatible only with functionalism as it appears in the philosophy of mind, but that persons are identified by functional properties.

> (1) There is psychological continuity if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if
> (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) it has not taken a 'branching' form. (5) Personal identity over time just consists in the holding of facts like (2) to (4).[20] (Parfit 1984, p. 207)

A person is a changing entity whose identity is determined only by her psychological continuity, independently from her personal traits and specific parts of her body. In accordance with the psychological continuity view, a person can change over a period of time in ways that would be characterized in everyday language as a complete change of personality or as a drastic bodily change. A person can lose or gain a number of qualities and capacities so long as she retains psychological continuity with her past self. So what about Alex Murphy, Will Caster, and Otto? Can we claim that they are still the same people after their hybridization? I believe that they are, as long as they are psychologically continuous with past Alex Murphy, Will Caster, and Otto. In Otto's case we can even claim that the retention of his personal identity was enabled by extension. The mnemonic device he uses helps him restore his memory capacity which is crucial for keeping up his psychological continuity.

Thus, even if these three cases can be considered as a confirmation of the retention of personal identity through hybridization, this confirmation is only contingent, and we can imagine cases in which we could not talk about such retention. These cases jeopardize versions of EM that are based solely on 1) and 2), and provide reasons for adding step 3) to the argument for EM. If we consider the original SR we will find one such example. It was claimed that the person inside the Chinese room, the written instructions, the database, and the scratchpads constitute a system. This system could not perform its main function—understand Chinese—without any of these component parts. According to our functionalist view of integration, we should consider these parts sufficiently integrated, and that they do indeed form a system. Also, according to the proponents of SR, we could identify mental states of this hybrid system (if it is capable of understanding, it has to be able to understand something in particular), but these mental states would not be attributable to the person in the room. Without the inclusion of step 3) in the argument for EM, this case could be counted as a counterexample to EM. There would be widely realized mental states without the extension of a particular mind, namely, the mind of the person in the room. Nevertheless, if we count psychological continuity as one of the criteria for EM, then there is a clear dissimilarity between Otto and the person from the Chinese room scenario. While narrow Otto and hybrid Otto are psychologically continuous, because there are chains of psychological

[20] (4) is added to safeguard the numerical identity of a person against cases where one person is duplicated by sci-fi devices such are teleporters.

connections between their mental states,[21] the person in the Chinese room and the whole system are strongly discontinuous in a psychological sense. Otto's occurrent beliefs produced all of the Otto-notebook system's dispositional beliefs, wishes and desires of narrow Otto still guide the behavior of hybrid Otto, hybrid Otto remembers some of the experiences of narrow Otto, etc. Contrary to this, the whole system's beliefs and the person-in-the-room's beliefs are not connected in a similar fashion. We can identify beliefs of the system by the answers to the questions posed in Chinese, but these answers and corresponding beliefs will not be appropriately psychologically connected with mental states of the person-in-the-room. Beliefs of these persons are completely disconnected in a psychological sense and may be thoroughly mutually inconsistent.[22]

Psychological continuity is similarly questioned in cases of distributed cognition, where multiple cognitive subjects are connected in ways that enable the emergence of new distributed cognitive processes (Hutchins 1995). If there are distributed systems which are sufficiently integrated, and attributable with intentional mental states, we could, arguably, treat those systems as persons.[23] Such distributed systems would be either psychologically continuous with more than one subject—realizing subjects with split-personalities non-identical to any of the individual persons in particular—or discontinuous with all individual persons that constitute them. We can also imagine cases in which a hybridization process impairs original cognitive capacities of a subject and creates psychological discontinuity. These cases would resemble that of Alex Murphy (if we imagine that Murphy did not start remembering his past life). They would include severe alterations of a person's normal cognitive functioning, such that result in a loss of access to her memories or a complete dissociation from them. In such cases instead of talking about the cognitive extension, we should rather categorize them as cases of creation of new hybrid cognitive subjects and destruction of original subjects. To conclude, we can distinguish at least three kinds of hybrid cognitive systems in which integrative functions and mental functions are realized, but in which there is no extension of the mind in the sense that there is no persistence of a unique person. They are hybrid systems that have original subjects as their parts but are

[21] Psychological connections between X and Y are established if X's mental states have been caused by Y's mental states, or if Y's mental states have been caused by X's mental states. Also, to secure psychological continuity, there has to be "enough connectedness" between X and Y, one psychological connection would not be sufficient. (Parfit 1984, p. 206) It is also important to note that not any causal connection between mental states would suffice. Causal relations have to be psychologically relevant, meaning that causes and effects should be semantically dependent. Intentional mental states are individuated by their semantic content, and relevant connections will be sensitive to this content (for instance, my belief that I should exercise more will be psychologically connected to my belief that exercising is healthy, my desire to be healthy, and so on). Mental states of the person-in-the-room and the whole system are not connected in this way. Narrow person's beliefs that she should organize symbols in some particular way are not psychologically connected to the wide person's belief that, for instance, the weather is nice.

[22] For more detailed arguments in favor of personal non-identity between the system and the person in the room see Maudlin (1989) and Cole (1991).

[23] I will say more about the plausibility of such systems in Sect. 6.3. While consideration of such possibility cannot be attributed to Hutchins himself, there are authors (Theiner 2014; Huebner 2014) who base their reasoning on the plausibility of full-blown group or collective minds, at least partly, on Hutchins's insights about cognitive distribution.

psychologically independent, systems with collective minds, and hybrid systems in which psychological continuity is broken by an impairment of cognitive capacities of the original subject.

6 Selves with wide realization

So far the question about the relevant personal identity was posed in such a way that presupposed the possibility of wide selves. Miyazono argued that PA leads us only to wide or hybrid selves (or to the claim that the Otto-notebook system is a proper subject of wide mental states), and my counterargument was that these wide selves can, at least sometimes, be considered as extended selves too (or that wide and narrow subjects can be identical). Nevertheless, there are authors who argue that wide selves cannot exist, even if there are wide mental states. In their view, the boundaries of a person, or a self, are not necessarily boundaries of her mind. Personal boundaries could be those of a human animal (or relevant parts of a human animal), and independent of her psychological boundaries. According to this view, in cases of wide mental states, or wide cognitive systems, we cannot conclude that persons are wide too (Olson 2011; Wilson 2004; Wilson and Lenart 2014; Baker 2009). Thus, it could be objected that without a proper defense of the view that there could be selves with wide realization, the defended claim about personal identity of narrow and wide subjects cannot stand on its own. In order to prevent this kind of objections to the view I am defending I will shortly address (a) Olson's Equivocation of 'Mind' Objection to extended selves (Olson 2011), and (b) Wilson and Lenart's view of extended personal identity without personal extension (Wilson 2004; Wilson and Lenart 2014); and offer (c) one possible positive view of widely realized selves.

6.1 Are extended selves possible?

In objecting to the claim about extended selves,[24] Olson (2011) starts from Clark and Chalmers's enthymematic argument: "Does extended mind implies an extended self? It seems so" (1998, p. 18). He then explicates its suppressed premise 2, and argues that the argument is "sophistical" as it rests on an equivocation in the use of the word "mind". He interprets the argument as follows:

1. Otto's mind extends as far as his mental states extend.
2. Otto = Otto's mind.
3. Thus, Otto extends as far as his mental states extend. (Olson 2011, p. 485)

In Olson's analysis the conclusion of the argument does not follow because in the first premise "mind" is used as an "abbreviation for Otto's mental states", and in the second

[24] Olson's objection can be generalized to all claims which presuppose the existence of wide selves. In that manner, his argument can be so rephrased to become an argument against the possibility that the Otto-notebook system is a subject of wide mental states (without specifications whether he is identical to Otto or not).

premise as "the subject of Otto's mental states" or "the thinking thing". (Olson 2011, p. 485)

We could assume that there could be a theory of mind that supports both readings, but Olson argues that there is none. In other words, according to Olson mind as a subject, which is the only true contender for a person, is not comprised of all of its mental states. In order to defend this claim, he introduces a strong metaphysical assumption. He claims that "mental internalism", or the metaphysical claim that all mental states are within the boundaries of the mind, is no truism and needs further argumentation, and supposes that "mental externalism" is a viable option. (2011, p. 485) Nevertheless, "mind externalism" can only be made tenable if we accept that realizers of mental properties or mental states can be wider than those of the mind whose properties they are, and thus reject the "flat" view of realization (Gillett 2002)—a widely held position according to which "the subjects of the realizer and realized properties are the same" (Rupert 2009, p. 62); or by assuming that mind, as a robust subject of mental states, is a shrunken mind—one that possesses only occurrent or conscious states or central control—to which dispositional states are external. It seems that Olson has something like this in mind. But this leaves one important question unanswered: Why would we call those external states *mental* in the first place if they are not part of the mind?[25] However controversial this assumption is, Olson continues to argue according to it and writes that we cannot simultaneously hold that the mind is the subject of mental states and that it extends with its states, because either the mind will be internally realized (as a fairly robust subject),[26] or it will be just a bundle of mental states, in a Humean sense, which cannot properly be deemed a subject of mental states.[27]

What is certain is that Clark and Chalmers do not have to suppose that minds are either "robust subjects" or simple sums of mental states. In other words, I believe that Olson presents us with a false dilemma—a dilemma which disappears when we accept a different picture of the mind, the one presented below (in subsection 6.3), and, arguably, tacitly endorsed in "The extended mind".

[25] If minds are to be conceived in this way, and external states are not really mental, then Olson's argument would be an argument against EM, and not an argument against the inference that extended selves are implied by EM. Other option for Olson would be to allow that there are two minds, one extended with all mental states, and one narrower which is a proper subject of these states, but this would be a strange view to defend.

[26] How robust Olson sees the mind as a subject is illustrated in his claim that if we reject the view that it is simply an organism that thinks this will lead us to "full-on substance dualism: a dualism of psychological and biological beings." (Olson 2011, p. 488) This way of thinking about minds could be one of the reasons why he rejects the idea that persons could be identical with their minds, and defends biological accounts of personhood further supported with arguments such as the Too Many Thinkers Argument.

[27] Interestingly enough, Clark and Chalmers claim quite the opposite, that shrinking of the mind would lead to a bundle view, and that taking the extended view gives a more (explanatory) robust view of the subject "Otto *himself* is best regarded as an extended system, a coupling of biological organism and external resources. To consistently resist this conclusion, we would have to shrink the self into a mere bundle of occurrent states, severely threatening its deep psychological continuity." (1998, p. 18).

6.2 Extended personal identity without extended persons

Wilson's view (2004) might, on the surface, look as a direct confirmation of Olson's claim that persons, as psychological beings, do not incorporate their external mental states. Namely, Wilson uses psychological criteria for personhood and personal identity, but at the same time rejects "mental internalism" by claiming that persons can have narrower boundaries than their mental states and cognitive systems of which they are a part (2004, p. 139). In fact, he is claiming that boundaries of bearers of mental states are "individual organisms—spatio-temporally bounded, relatively cohesive, unified entities that are continuous across space and time." (2004, p. 143) As described, this position is in direct conflict with Clark and Chalmers's claim about extended selves, and Miyazono's and mine assumption that wide minds imply wide selves, as it allows for extended or wide mental states and cognitive systems, but not for extended or wide persons.

Without further explanations about Wilson's position we may wrongly infer that the reasoning behind these claims is similar to that of Olson, in a sense that not all mental states are constitutive of personhood if the person is taken to be the subject of mental states (for instance, person's dispositional beliefs, or any state which is not followed by the experience of "I"). But it soon becomes clear that Wilson's preference for "mental externalism" is not based on such reasoning. In (2014) Wilson and Lenart argue that even personal identity can be widely realized, as some of the memories constitutive of personal identity can be widely, even collectively realized, but at the same time they argue that this view does not imply that the boundaries of persons are also extended. This may leave us puzzled. There are intrinsic constitutive properties of a person, namely those properties which make this person a person that she is, which nevertheless do not fall within her boundaries. So what leads Wilson and Lenart to defend such a view?

Answers are to be found in Wilson's view of realization, and his specific version of extended mind thesis.[28] Wilson's extended mind position is radically different from that of Clark and Chalmers, and it is based on vastly different arguments. Wilson's thesis about the wide realization of mental states draws on semantic externalism or anti-individualism of Putnam (1975) and Burge (1979), and context sensitivity. He argues in favor of wide realization of mental states with wide contents because relevant environmental factors are unavoidable in their individuation.[29] He combines the view of semantic externalism with a specific view of realization that is based on sufficiency and not constitutivity. According to Wilson, we can differentiate between core realizers, which play crucial causal roles of the relevant property identified by functionalism, and total realizers, which are metaphysically sufficient for some higher-level property (2004, pp. 107–108). Wilson's claim about wide realization is concerned with

[28] Wilson and Lenart (2014) base their claim that extended mind does not imply extended self on Wilson's work (2004), and briefly quote that "characterization of wide realizations preserves the idea that properties with such realizations are still properties of individual subjects" (Wilson 2004, p. 141; Wilson and Lenart 2014, p. 433). This is why I continue with Wilson's justification of this claim from (2004).

[29] In Putnam's Twin Earth scenario, Oscar from Earth, and Oscar* from Twin Earth are in different mental states, although core realizations of their mental states are identical and within the boundaries of their bodies.

total realizations only, and, in his view, core realizations are entity-bounded. Collective wide realizations of mental states, even those strongly constitutive of personal identity, are such because we can talk about "folk psychological systems" which involve social relations between individuals (inspired by views of semantic deference in linguistic social practices, 2004, p. 113), but they are to be seen as total and not core realizations. Thus, "mental externalism" holds only if we take realization in a sufficiency sense and look at total realizations of mental states, while "mental internalism" still holds for core (constitutive) realizations.

Thus, there is no direct conflict between Wilson and Clark and Chalmers with respect to the boundaries of the self. Clark and Chalmers argue in favor of wide core realizations (realizations that are identified by crucial causal roles), and according to Wilson they are entity-bounded so they must be a part of a wide entity if they exist. The real conflict between these two positions might be about the possibility of wide core realizations, but with that question we already dealt in detail.

6.3 Wide cognitive systems and wide selves

By now it is clear that we can present a simple defense of extended or wide selves by insisting on "mental internalism". If persons are simply subjects of mental states, and we adopt the "flat" or constitutive view of realization, then if mental properties are widely realized, so is the subject of these properties. Still, there are further arguments that persons are not simply subjects of mental states, which aim to show that persons are organism-bound, or that they are special parts of minds, and which therefore ask for a more robust positive view of extended selves. There are two kinds of arguments in favor of organism-bounded persons: those that point out that there are no other proper candidates for "spatio-temporally bounded, relatively cohesive, unified entities" (Wilson 2004, p. 143; Rupert 2009), and those that insist on biological criteria of personhood (Snowdon 1990; Olson 1997). I have tried to defend a view according to which hybrid cognitive systems may play a role of relatively cohesive, unified entities, based on integration conditions, so in the remainder of this section I will focus on answering why biological criteria do not suffice for identification of persons, and why we should not consider persons to be some special parts of minds.

One of the main arguments in favor of biological or animalist approach to persons, and against the psychological account, is the Too Many Thinkers Argument (Snowdon 1990; Olson 1997). According to it, we can claim that the human animal writing this sentence and sitting in my chair is thinking, so that if I—who thinks—is something different from that animal, then there are two, or too many, thinkers. This leads to a range of problems, and one of the most severe is certainly that I could never know who I am, whether I am a biological or a psychological being, because I cannot be both.[30] Shoemaker (1999) turns the objection against those who defend it by claiming that if we follow this course of reasoning then, even if we deny that there is a psychological

[30] If there would be two or more persons sharing mental states and their body, there would have to be double or multiple attributions of agency in particular cases. Who would be held responsible for paying a bill in a restaurant, one or the other person, or both? Mackie says "that it just seems crazy to suppose that there are two psychological lives going on where I am now". (1999, p. 375).

person, there would still be too many thinkers: the animal's body (as different from the living animal), perhaps a brain of the animal which thinks, etc. What Shoemaker tries to emphasize is that physical bases of different higher-level properties can contingently coincide (as in the case of the body of an animal, an animal, and a psychological person if she has a narrow realization), but that this does not make these properties identical. Identifying the property of being a person with the property of being a human animal misses the right level of description, as it leads to scientific inexplicability of agency and why we attribute mental states to such entities in the first place.

Psychological accounts of persons have an explanatory advantage over biological accounts because *being a person* is most readily identified as a complex psychological property of the thinking entity capable of attributing intentional states to others and itself, and capable of telling a story about itself. Identifying the boundaries of the physical system which instantiates these properties, as different from those of a human animal, does not have to be "full-on substance dualism", as Olson claims (2011, p. 488), but we can rather see psychological entities as any other kind of entities postulated by special sciences. The question is then: What kind of a physical system realizes the higher-level property *being a person*?

Here I follow naturalistic approach inspired by Daniel Dennett's work (1987, 1992, 1996) with a more realistic turn. Persons are nothing but a special kind of minds. They are higher order intentional systems capable of attributing intentional states to others and to themselves. Arguably (see Carruthers 2006), such capabilities of these systems are enabled by the use of language. The crucial question is what is this thing that is capable of such actions? Does it have to be some "central meaner" which understands those intentional states thanks to the existence of original intentionality? The answer is "no". With Turing's definition of computation which allows for mechanical manipulation of symbols with semantic content without an entity which understands the symbols, and with naturalistic views of intentionality which do not evoke the difference between original[31] and derived intentionality, the path was paved to thinking that minds are not some things that people have and other creatures do not, but that there are creatures which are minded, meaning that they cope with the environment in more flexible ways thanks to representing its features. In this picture, cognition, mentality, minds, and persons are to be explained by mechanisms which are a product of both biological and cultural evolution. According to it, there are many kinds of minds, animal minds of different structures, or different kinds of intentional systems whose representational states enable them to cope with changes in the environment in ways which cannot be properly described as fixed pattern behaviors. Although we can instrumentally attribute full-blown intentional mental states to all intentional systems in explaining their behavior,[32] that does not mean that we can or should attribute them realistically to all of them. We can differentiate between minimal and maximal minds (and other kinds in between), according to the kinds of representational states and processes they instantiate. Only those to which we can realistically attribute struc-

[31] Original intentionality is also derived in a sense that it derives its meaning from Mother Nature (Dennett 1987).

[32] I can attribute different desires and beliefs to my cat in explaining her behavior, but that does not mean that she has, or is capable of having propositional attitudes.

tured propositional attitudes are maximal—we call them "selves" or "persons"—and others will have different kinds of less sophisticated mental states which do not have to be propositionally structured (Huebner 2014). What counts as a suitable level for identifying the physical basis of minds or persons is the level of cognitive architecture that allows for attribution of psychological properties. This architecture, arguably, "consists of relatively independent subsystems, which each process a narrow range of information, and which can be coordinated and interfaced to facilitate skillful coping with environmental contingencies" (Huebner 2014, p. 199) This kind of functionalist approach will make a reference to different kinds of biological subsystems, created in different times of evolutionary history, which process information, and in that sense it "explains why the body matters without making the body matter mysteriously" (Clark 2008, p. 206), as biological account does. Also, focusing on the functional organization will allow for multiple realizability of appropriate structures, and sometimes cognitive systems will be constituted of both biological and non-biological parts. Not less importantly, this inclusion of lower level subsystems as essential in cognitive architecture is the reason why we cannot focus only on the very top of the hierarchy of cognitive mechanisms in search of the relevant physical realization, and why it is fruitless to search for some core structure, some specific realizer of the self. As Clark writes:

> There is no self, if by self we mean some central cognitive essence that makes me who and what I am. In its place there is just the "soft self": a rough-and-tumble control-sharing coalition of processes – some neural, some bodily, some technological – and an ongoing drive to tell a story, to paint a picture in which "I" am the central player. (Clark 2003, p. 138)

Once we allow that so described cognitive systems are selves—systems with appropriate information processing subsystems, which behave in a flexible way, and to which we can ascribe propositionally structured intentional mental states—and that these kinds of systems can have non-biological parts, then we should also allow for the existence of extended or wide selves. Also, it is important to note that once we consider a particular extended cognitive system as a self, we have to reject the possibility that there is simultaneously a narrower system (a biological organism, or a brain), which is psychologically continuous with the same past person as this hybrid system. In other words, once Otto becomes Otto-notebook he ceases to be biological Otto. In the opposite case, the Too Many Thinkers problem would occur just as in the case that my organism minus a brain part, which contains some of my memories, instantiates a second person. There would be no clear psychological or physical criteria to distinguish such persons, and almost all of their mental states and actions would be attributable to both of them.[33]

The remaining question is whether we can attribute personhood to group minds if we take persons to be only those cognitive systems which can be realistically attributed with mental states such as beliefs, desires, etc. Bryce Huebner in *Macrocognition* takes this question seriously and investigates a number of candidates for group minds—from

[33] This presents a second indirect argument against Miyazono, because if he grants that the Otto-notebook system is a self, and that it is not Otto, then Otto will perish in the process of hybridization.

Hutchins's example of ship navigation system to scientific collaborative groups (such as the one existing at CERN). His conclusion is that, while it is common in the literature on collective intentionality and collective agency to attribute propositional attitudes to groups, this can usually be justified only instrumentally. He admits that the best candidates for true collective minds, such as Hutchins's naval example of "fix cycle" system, are indeed minds but not maximal. On the other hand, he claims that he sees no "obvious reason to deny the status of a maximal mentality to at least some of the outputs at CERN, but establishing that they do would require a great deal more anthropological research." (Huebner 2014, p. 255) Thus, if we take Huebner's careful and insightful analysis of collective mentality and its conclusions as correct, we should accept that collective minds, and collective persons, are plausible, though extremely rare.

7 Concluding remarks

Let us consider the argument for EM as reconstructed by Gertler once more:

(1)"What makes some information count as a [standing] belief is the role it plays" (1998, p. 14).
(2)"The information in the notebook functions just like [that is, it plays the same role as] the information constituting an ordinary non-occurrent belief". (1998, p. 13)
(3)The information in Otto's notebook counts as standing beliefs. (from (1) and (2))
(4)Otto's standing beliefs are part of his mind.
(5)The information in Otto's notebook is part of Otto's mind. (from (3) and (4))
(6)Otto's notebook belongs to the world external to Otto's skin, i.e., the 'external' world.
(7)The mind extends into the world. (from (5) and (6)) (Gertler 2007, p. 193)

I accept (1) and (2). Premise (2) can be given a wide reading where "functions just like" refers to both kinds of functional roles, integrative and mental, in order to account for the request of step 1). But to avoid confusion I believe it is better to add an additional premise:

(2*) Otto and the notebook form a system. (based on integration considerations)

Further, I agree with Miyazono that (5) does not follow from (3) and (4), because it could have been the case that the hybrid system is not identical with Otto. I tried to show why I believe that Miyazono's argument is flawed. He argued that in every case an extended mental state should be attributed to a hybrid system and not to an original subject, based on his interpretation of SR. As it is shown, non-identity of a hybrid system and its central part does not follow from the mere fact that there is a hybrid system. Non-trivial personal non-identity can only follow from the fact of psychological discontinuity. And for this reason I suggest the inclusion of the following premise in the argument for EM:

(3*) Otto is identical to the Otto-notebook system and all mental states of the system can be attributed to him. (based on psychological continuity)

Justification of all three steps of the argument are functionalist, but employ different functional theories, namely, a functional description of integration, functionalism about the mental, and functional theory of personal identity. None of these theories assume a specific kind of matter or specific components of cognitive systems. Thus, I believe that I have alleviated the "great difficulty in conceiving even a single case where, according to some sort of functionalism, OEB (or a corresponding claim) is true." (Miyazono 2017, p. 3540) This difficulty in conceiving a functionalist EM was a product of great expectations: expectations that one set of functional criteria can give us conditions of integration, cognitiveness or mentality and personal identity. This was never intended. PA was offered as a crucial step in arguing for EM under the assumptions of integration and identity, which were not a priori questionable. Under the attack, the identity assumption in Otto's case prevailed, as the application of SR to EM did not succeed. SR in its original form did show that the hybrid system is different from the person in the room, because of the mutual independence of this person's mental states and the hybrid system's mental states, but the same does not apply to Otto; and Mark and his book were just a red herring disguised as a cognitive system. Even though the objection was unsuccessful, it pointed out one important aspect of the argument for EM. This argument does rest on the assumption of personal identity, which has to be justified. But not solely to defeat SR, but to act as a point of demarcation between the cases of extended minds and those in which minds are distributed or discontinuous with persons that entered the process of hybridization. And while functionalism (in its narrow sense) does not entail EM without additional assumptions (which was never disputed), it certainly does not mean that functionalism is incompatible with EM or that "extended mental states are metaphysically impossible according to functionalist theories of mind." (Miyazono 2017, p. 3540)

Acknowledgements I thank Brian Leahy, Andrej Jandric and Vojislav Bozickovic for carefully reading and commenting on previous versions of the manuscript. Their suggestions and observations helped to greatly improve it. I would also like to thank two anonymous referees for helpful comments and valuable suggestions that shaped the final version of this article.

Funding This research was supported by the Ministry of education, science and technological development of the Republic of Serbia under the project "Dynamical systems in nature and society: philosophical and empirical aspects" (179041).

References

Adams, F. R., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology, 14*(1), 43–64.
Adams, F. R., & Aizawa, K. (2008). *The bounds of cognition*. Oxford: Wiley-Blackwell.
Baker, L. R. (2009). Persons and the extended-mind thesis. *Zygon, 44*(3), 642–658.
Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy, 4*(1), 73–122.
Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought*. Oxford: Oxford University Press.
Clark, A. (2003). *Natural-born cyborgs: Minds, technologies and the future of human intelligence*. Oxford: Oxford University Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.
Clark, A. (2010). Memento's revenge: The extended mind extended. In Richard Menary (Ed.), *The extended mind* (pp. 43–66). Cambridge, MA: MIT Press.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58,* 7–19.
Cole, D. J. (1991). Artificial intelligence and personal identity. *Synthese, 88,* 399–417.
Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown & Company.
Dennett, D. C. (1992). The self as a center of narrative gravity. In F. S. Kessel, P. M. Cole, & D. L. Johnson (Eds.), *Self and consciousness: Multiple perspectives* (pp. 4–237). New Jersey: Lawrence Erlbaum.
Dennett, D. C. (1996). *Kinds of minds*. New York: Basic Books.
Donald, M. (1991). *Origins of the modern mind*. Cambridge, MA: Harvard University Press.
Drayson, Z. (2010). Extended cognition and the metaphysics of mind. *Cognitive Systems Research, 11*(4), 367–377.
Gertler, B. (2007). Overextending the mind? In B. Gertler & L. Shapiro (Eds.), *Arguing about the mind* (pp. 192–206). Abingdon: Routledge.
Gillett, C. (2002). The dimensions of realization: A critique of the standard view. *Analysis, 62*(4), 316–323.
Heersmink, R. (2017). Distributed cognition and distributed morality: Agency, artifacts and systems. *Science and Engineering Ethics, 23*(2), 431–448.
Heidegger, M. (1927/1961). *Being and time,* trans. J. Macquarrie & E. Robinson. New York: Harper & Row.
Huebner, B. (2014). *Macrocognition: A theory of distributed minds and collective intentionality*. New York: Oxford University Press.
Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy, 50*(3), 249–258.
Lewis, D. (1980). Mad pain and Martian pain. In N. Block (Ed.), *Readings in the philosophy of psychology* (pp. 216–222). Cambridge, MA: Harvard University Press.
Locke, J. (1689/1975). An essay concerning human understanding. In P. H. Nidditch (Ed.), *The clarendon edition of the works of john locke*. Oxford: Oxford University Press.
Mackie, D. (1999). Animalism versus lockeanism: No contest. *Philosophical Quarterly, 50*(196), 369–376.
Maudlin, T. (1989). Computation and consciousness. *Journal of Philosophy, 86,* 407–432.
Menary, R. (2007). *Cognitive integration: Mind and cognition unbounded*. Basingstoke: Palgrave-Macmillan.
Miyazono, K. (2017). Does functionalism entail extended mind? *Synthese, 194*(9), 3523–3541.
Olson, E. T. (1997). *The human animal: Personal identity without psychology*. Oxford: Oxford University Press.
Olson, E. T. (2011). The extended self. *Minds and Machines, 21*(4), 481–495.
Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.
Putnam, H. (1960). Minds and machines. In Sidney Hook (Ed.), *Journal of symbolic logic* (pp. 57–80). New York: New York University Press.
Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science, 7,* 131–193.
Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy, 101,* 389–428.
Rupert, R. D. (2009). *Cognitive systems and the extended mind*. New York: Oxford University Press.
Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences, 3,* 417–457.
Shapiro, L. A. (2004). *The mind incarnate*. Cambridge, MA: A Bradford Book.
Shoemaker, S. (1999). Self, body, and coincidence. *Proceedings of the Aristotelian Society, 73*(73), 287–306.
Snowdon, P. F. (1990). Persons, animals, and ourselves. In Christopher Gill (Ed.), *The person and the human mind: Issues in ancient and modern philosophy*. Oxford: Oxford University Press.
Sprevak, M. (2009). Extended cognition and functionalism. *Journal of Philosophy, 106*(9), 503–527.
Sutton, J. (2010). Exograms, interdisciplinarity and the cognitive life of things. In R. Menary (Ed.), *The extended mind*. Cambridge, MA: MIT Press.
Theiner, Georg. (2014). A beginner's guide to group minds. In Kallestrup Jesper & Sprevak Mark (Eds.), *New waves in philosophy of mind* (pp. 301–322). Palgrave-Macmillan: Basingstoke.
Walter, S. (2010). Cognitive extension: The parity argument, functionalism, and the mark of the cognitive. *Synthese, 177,* 285–300.

Wheeler, M. (2010). In defence of extended functionalism. In R. Menary (Ed.), *The extended mind*. Cambridge, MA: MIT Press.

Wilson, R. A. (2004). *Boundaries of the mind: The individual in the fragile sciences: Cognition*. New York: Cambridge University Press.

Wilson, R. A., & Lenart, B. (2014). Extended mind and identity. In J. Clausen & N. Levy (Eds.), *Handbook of neuroethics* (pp. 423–439). NewYork: Springer.