SITUATED COGNITION AND ITS CRITICS: RECENT DEVELOPMENTS

EDITED BY: Albert Newen, Beate Krickel, Achim Stephan and Leon De Bruin
PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-645-6
DOI 10.3389/978-2-88971-645-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

SITUATED COGNITION AND ITS CRITICS: RECENT DEVELOPMENTS

Topic Editors:
Albert Newen, Ruhr University Bochum, Germany
Beate Krickel, Technical University of Berlin, Germany
Achim Stephan, University of Osnabrück, Germany
Leon De Bruin, Radboud University Nijmegen, Netherlands

Citation: Newen, A., Krickel, B., Stephan, A., De Bruin, L., eds. (2021). Situated Cognition and Its Critics: Recent Developments. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-645-6

Table of Contents

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 09 April 2020
doi: 10.3389/fpsyg.2020.00674

From Notebooks to Institutions: The Case for Symbiotic Cognition

*Marc Slors**

Faculty of Philosophy, Theology and Religious Studies, Radboud University, Nijmegen, Netherlands

Cognition is claimed to be extended by a wide array of items, ranging from notebooks to social institutions. Although the connection between individuals and these items is usually referred to as "coupling," the difference between notebooks and social institutions is so vast that the meaning of "coupling" is bound to be different in each of these cases. In this paper I argue that the radical difference between "artifact-extended cognition" and "socially extended cognition" is not sufficiently highlighted in the literature. I argue that there are two different senses of "cognitive extension" at play, that I shall label, respectively, "implementation extension" and "impact extension." Whereas implementation extension is a causal-functional notion, impact-extension hinges on social normativity that is connected with organization and action coordination. I will argue that the two kinds of cognitive extension are different enough to warrant separate labels. Because the most salient form of social extension of cognition involves the reciprocal co-constitution of cognitive capacities, I will propose to set it apart from other types of extended cognition by using the label "symbiotic cognition."

Keywords: extended cognition, socially extended cognition, cognitive integration, distributed cognition, symbiotic cognition

OPEN ACCESS

Edited by:
Achim Stephan,
Osnabrück University, Germany

Reviewed by:
Raoul Gervais,
University of Antwerp, Belgium
Santiago Arango-Munoz,
University of Antioquia, Colombia

***Correspondence:**
Marc Slors
m.slors@ftr.ru.nl

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *05 February 2020*
Accepted: *19 March 2020*
Published: *09 April 2020*

Citation:
Slors M (2020) From Notebooks to Institutions: The Case for Symbiotic Cognition. Front. Psychol. 11:674. doi: 10.3389/fpsyg.2020.00674

INTRODUCTION

In the literature on extended, integrated and distributed cognition, human cognitive systems are said to be coupled with and enhanced by a large number of rather diverse items, ranging from simple notebooks and abacuses (Clark and Chalmers, 1998), via complete physical environments such as theater set-ups (Tribble, 2005; Clark, 2008; Sutton, 2010) to language (Clark, 2008) and social institutions such as legal systems (Fuchs and De Jaegher, 2009; Gallagher and Crisafi, 2009; De Jaegher et al., 2010; Gallagher, 2013; Gallagher et al., 2019). This range is so wide, and the difference between e.g., a notebook and a social institution so immense, that it seems unlikely that people are connected with these items in basically the same way. In this respect it doesn't matter whether we speak of cognitive integration (Menary, 2007, 2010), "distributed cognition" (Hutchins, 1995; Hutto and Myin, 2017) or of cognitive extension (Clark and Chalmers, 1998; Clark, 2008; Gallagher, 2013). The point is that just saying that we are "functionally integrated" (Heersmink, 2015) with items or "causally coupled" with them is bound to sweep an important difference under the carpet when these items are so radically different. The aim of this paper is to characterize the difference between the way our cognition is extended by and/or integrated with items such as notebooks, abacuses, and smart phones on the one hand—which I will call artifact-extended cognition—and items such as social institutions, language, and cultural conventions—which is known as socially extended cognition—on the other. I will argue that the difference is significant enough for the latter kind of extension/integration to warrants its own separate label, for which I will propose the term

"symbiotic cognition." In order not to complicate the discussion unnecessarily, I will concentrate on the literature on extended cognition, except for the last section of this paper.

The paper is set-up as follows. In the next section I will introduce the notion of extended cognition and highlight the difference between artifact-extended cognition and socially extended cognition. In the section "The Problem of Cognitive Bloat," I will briefly discuss the problem of cognitive bloat such as this has first been proposed as an argument against the early varieties of cognitive extension. I will argue that if socially extended cognition is indeed modeled on artifact-extended cognition, it falls prey to this problem in such a blatant way that it is clear that we must understand socially extended cognition differently. In the section "Implementation-Extension and Impact-Extension," I will propose a characterization of the difference between artifact-extended cognition and socially extended cognition. I will argue that cognition can be considered to be extended in different ways. Whereas artifact-extended cognition extends cognitive processes by extending the *implementation base* of these processes, socially extended cognition alters the nature and hence extends the *impact* of cognitive engagements with the world by embedding them in social practices of coordinated behavior. When we interpret socially extended cognition as an instance of impact-extension and not as implementation-extension, the problem of cognitive bloat disappears.

In the section "Causality, Coordination, and Reciprocal Cognitive Dependency," I will defend and elaborate on the distinction between "implementation-extension" and "impact-extension" by arguing that, crucially, the chain of items causally linked to a person whose cognition is socially extended involves other human beings—other cognitive systems. On the one hand, this introduces social normativity into the extended system, which is absent in artifact-extended cognition. On the other it introduces the idea of reciprocal cognitive dependency between people. I will propose the label "symbiotic cognition" for networks of mutually dependent cognitive systems. In the section "Cognitive Symbiosis, Weak and Strong," I will define the notion of symbiotic cognition. I will allow for the possibility of socially extended cognition that is not symbiotic cognition, and will distinguish between weak forms of symbiotic cognition, that do not require social institutions, and strong forms that do. In the section "Symbiotic Cognition, Cognitive Integration and Distributed Cognition," I will compare the idea of symbiotic cognition with integrated cognition (Menary, 2007, 2010, 2013) and distributed cognition (Hutchins, 1995; Hutto and Myin, 2017). I will argue that although some elements of symbiotic cognition surface in these views, the essential contrast between artifact- and social extension is still ignored by both.

ARTIFACT-EXTENDED AND SOCIALLY EXTENDED COGNITION

The idea that human cognitive systems are in fact extended by items outside our brains and bodies has been developed and defended by many philosophers for over two decades now. Disregarding precursors, the idea that started the debate on extended cognition—then labeled "active externalism" (Hurley, 2010)—was based on the so-called parity principle: "If, as we confront some task, a part of the world functions as a process which were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (. . .) part of the cognitive process." (Clark and Chalmers, 1998, 8) The (in)famous and widely discussed Otto and Inga example exemplifies this principle: Inga wants to visit the MoMa in New York and remembers that it is on 11 West 53rd Street. Otto has early onset Alzheimer. Instead of relying on information storage in his head, he uses a notebook that he always carries with him. When he wants to visit the MoMa he consults his notebook to find the address. Because the system consisting of Otto-and-notebook is functionally equivalent to Inga, Clark and Chalmers claim that the notebook is as much part of Otto's mind as the memory storage region of Inga's brain is part of hers.

Brain-chauvinists think there is a relevant difference. On their view, Otto does not remember the MoMa address. Rather, he believes the address is in the notebook, perceives the contents of the notebook and forms a new belief about the address. On this reconstruction all the mental work is done in Otto's head, not outside it. The response to this "Otto two-step" (Clark, 2010, 46) is easy enough to imagine: if Otto believes the required information is stored in the notebook and retrieved by perceiving the notebooks contents, then why not say that Inga believes the information she seeks is stored in her brain and she introspects the retrieval of that information, forming a new belief about the address of the MoMa? If we think this reconstruction of Inga's remembering is contrived then why is a similar reconstruction of Otto's mental processing not also contrived? The point is that the Otto two-step works only if we are already inclined toward brain-chauvinism.

Some philosophers have argued that we have reason to be chauvinist (Adams and Aizawa, 2001, 2008, 2010). These reasons involve an appeal to "non-derived mental content." But that is a controversial notion (Dennett, 1978, 1987; Clark, 2010; Hutto and Myin, 2013, 2017). I shall not go into the debate on non-derived content, because it is associated mostly with the "first-wave" extended cognition theories based on the parity principle.[1] The wider variety of items that our cognition is said to be coupled with, which is the main topic of this paper, stems mainly from the second wave of extended mind theories. These are not based on the parity principle, but on the complementarity principle (Sutton, 2010): items external to our brains and bodies can contribute to cognition, not because they structurally resemble processes that also occur inside the brain, but because they complement brain processes and by doing so allow for new cognitive possibilities. With the first-wave, parity principle-based extended cognition, it is possible to ask whether an extended process that resembles a brain process is as good as "the real thing" (and those who believe in non-derived content think it

[1] A second reason to leave this discussion for what it is, is because the arguments for brain chauvinism are directed against what I will later call "implementation extension." The notion of "impact extension" which I will argue characterizes socially extended cognition is not directly susceptible, I believe, to Adams and Aizawa's critique.

isn't). With the second-wave, complementarity principle-based type of extended cognition, this issue does not arise if the cognitive capacities that emerge when an embodied brain is coupled with external devices does not have a merely brain-based equivalent. The issue in such cases is simply whether we are or should be inclined to think of the emerging capacity as a cognitive one.

A prominent example of second-wave extended cognition is the idea that external symbol systems such as written language and numbers extend our cognitive capacities. In Menary's words "[t]he surrounding linguistic environment contains reliable structures, speech and text, that are available as cognitive resources to be coupled with. Our ability to reliably couple with this ever-present environment constitutes human cognition and thought" (Menary, 2010, 8). Clark agrees by emphasizing that "linguistic tools enable us to deliberatively and systematically sculpt and modify our own processes of selective attention." (Clark, 2008, 48) Physical symbols, whether written or spoken, bring about capacities for thought, communication and numeracy that are literally unthinkable without them.

The example of language and number-symbols systems is a good stepping stone to the idea of socially extended cognition (Fuchs and De Jaegher, 2009; Gallagher and Crisafi, 2009; De Jaegher et al., 2010; Gallagher, 2013; Gallagher et al., 2019). Just like written language and numbers can extend our cognitive range, the idea behind this position is that some social institutions can do that just as well. I will focus on Shaun Gallagher's exposition and defense of this idea, because it is the most elaborate version available. Gallagher argues for "a liberal, and specifically social extension of the extended mind hypothesis." He "appeal[s] to social practices and institutions that are what we might call "mental institutions" (Gallagher and Crisafi, 2009), in the sense that they are not only institutions with which we accomplish certain cognitive processes, but also are such that without them such cognitive processes would no longer exist." (Gallagher, 2013, p. 6) Examples he uses include legal systems, educational systems and museums, cultural conventions and even the market economy (Gallagher et al., 2019). The idea is that such institutions extend our cognitive capabilities considerably and that this should count as a form of extended cognition.

Our legal system, for example, enables an array of thoughts and actions that would not merely be impossible, but would not even be intelligible without the concept and procedural routines associated with the law. A helpful example that is often used by Gallagher is the practice of formalizing an agreement between two people by signing a contract:

> A contract or legal agreement (. . .) is in some real sense an expression of several minds externalized and extended into the world, instantiating in external memory an agreed-upon decision, adding to a system of rights and laws that transcend the particularities of any individual's mind. Contracts are institutions that embody conceptual schemas that, in turn, contribute to and shape our cognitive processes (Gallagher, 2013, p. 6).

The point I wish to make in this section is that somewhere along the way in ascending from notebooks as possible cognitive extensions to socio-cultural institutions, a crucial distinction is ignored. The connection between individuals and the items their cognition is extended with is described as more or less similar—it is described as "coupling." What coupling entails must depend on what we couple with. Hence, in order to maintain similarity throughout the ascent from notebooks to institutions, items that are said to extend our cognition are very often (but not always) described as physical objects. Language, for example, is described as a set of physical symbols. But apart from involving a set of physical symbols, language is also a social *practice*. It is not just scribbles and sounds, but also the way we use these in social interactions.

This certainly goes for social institutions too. Legal systems involve courtrooms, togas, and in some countries wigs. But they also involve rules, conventions and practices. A contract is an externalized memory not just because of its physical properties but mainly because of the way these pieces of paper (or bunches of bits) function in legal practice. Gallagher acknowledges that cognition can also be extended by institutions that are less formal and reinforced, such as practices involving cultural conventions:

> In solving a problem like keeping my cattle in my pasture, my bodily manipulations of a set of wooden poles and wire are not necessarily part of the cognitive process; but my engagement with the particular local custom/practice of solving this problem with a fence (and even a specific kind of fence) is a cognitive part of the problem solving. In such cases, cultural practices, local know-how in the form of established practices, etc., in either formal or informal ways, enter into and shape the thinking process. Without such cultural practices, rules, norms, etc. our thinking – our cognitive processes – would be different (Gallagher, 2013, p. 10).

Interestingly, the difference between physical objects and practices is not seen as an obstacle for claiming that coupling is basically similar when we move from notebooks to institutions:

> Just as a notebook or a hand-held piece of technology may be viewed as affording a way to enhance or extend our mental possibilities, so our encounters with others, especially in the context of various institutional procedures and social practices may offer structures that support and extend our cognitive abilities (Gallagher, 2013, p.4).

Let us call cognition that is extended by physical objects "artifact-extended cognition." The (admittedly rhetorical) question I would like to pose is whether Gallagher, Clark and Menary are correct (on some interpretations of their views) in assuming that socially extended cognition is really continuous with artifact-extended cognition. Are coupling with artifacts and coupling with practices really similar enough to warrant the use of the same label—extended cognition—in both instances?

THE PROBLEM OF COGNITIVE BLOAT

In order to make a beginning with driving a wedge between artifact-extended cognition and socially extended cognition, it is useful to look at what is known as the problem of cognitive bloat (Rupert, 2004). This is the problem that if we allow notebooks and smart phones to co-constitute our cognitive processes, we may have to include many other things too, in

which case we are likely to end up with cognitive processes that are so wide and scattered that it is counterintuitive to think of them as processes of a single person. I will argue that this problem is not alike for artifact-extended cognition and socially extended cognition.

From the perspective of artifact-extended cognition, Otto-and-notebook-style, the response to the threat of cognitive bloat is to tighten the constraints on what counts as co-constituents of cognition. Clark proposes four extra constraints:

(1) That the resource be reliably available and typically invoked. (Otto always carries the notebook and won't answer that he "doesn't know" until after he has consulted it).
(2) That any information thus retrieved be more or less automatically endorsed. It should not usually be subject to critical scrutiny (e.g., unlike the opinions of other people). It should be deemed about as trustworthy as something retrieved clearly from biological memory.
(3) That information contained in the resource should be easily accessible as and when required.
(4) That the information in the notebook has been consciously endorsed at some point in the past and indeed is there as a consequence of this endorsement (Clark, 2008, 79).

This does limit the possible candidate artifacts that may be said to extend cognition considerably. Arguably, the remaining problem is a matter of intuition. It is surely the case that even with these extra criteria our extended minds are bigger and more scattered than traditional brain-based or neo-Cartesian intuitions would make them out to be. But they are not so large and scattered that it is incoherent to think of them as single cognitive systems.

One of the reasons for this is that the external items we are said to be coupled with are not themselves coupled with still further structures in ways that satisfy 1–4. But this is exactly the problem with socially extended cognition. If we are coupled with social institutions, we are coupled with structures that are constituted, among other things, by (very many) other human beings. These human beings are themselves coupled with further structures in the same way we are coupled with them. And this makes the cognitive system implausibly large and scattered—if we are able to draw boundaries at all. For this reason, even philosophers who are sympathetic to the idea that human cognition involves massive coupling with our external niches are reluctant to think of social institutions as co-constituents of our cognitive systems (Huebner, 2013; Menary, 2013). According to them it is much more plausible to think of social institutions as the enabling conditions for cognitive abilities such as being able to sign contracts, speaking a language, or using cultural conventions.

The point I wish to make here is not that socially extended cognition clearly falls prey to the problem of cognitive bloat. Rather, the point is that (i) it would fall prey to the problem of cognitive bloat if socially extended cognition is a proposal that is modeled completely on the idea of artifact-extended cognition, and (ii) if it is interpreted in this way it falls prey to the problem of cognitive bloat so obviously and blatantly that it seems unlikely that socially extended cognition is intended to be modeled completely on artifact-extended cognition.

Gallagher is ambivalent here. On the one hand he does present socially extended cognition as a proposal that is somehow derived from the idea of artifact-extended cognition (see the last quote of the previous section "Artifact-Extended and Socially Extended Cognition"). On the other hand, however, he distances himself from Clark's functionalism and the way Clark deals with the problem of cognitive bloat. Tightening the restrictions on what counts as proper cognitive extension in the way Clark does, emphasizes the idea that the brain is still the central hub of any cognitive system, however extended this system is. And it is precisely such brain-centeredness that Gallagher wishes to overcome with the idea of socially extended cognition. But now the question arises: how is avoiding brain-centeredness and including social practices and institutions in the list of co-constituents of our cognitive processes going to help sidestep the problem of cognitive bloat?

I believe the answer here is to distance the idea of socially extended cognition even more from the idea of artifact-extended cognition than Gallagher does.

IMPLEMENTATION-EXTENSION AND IMPACT-EXTENSION

To say that cognition is extended is to say that items external to our brains and bodies expand our cognitive repertoire in such a way that they can somehow be said to co-constitute the "mechanisms"[2] of the cognitive system responsible for that repertoire. Differently put: some of the cognitive work in our interactions with the world has to be performed by items external to our brains and bodies. I believe there are different ways in which these descriptions can be made more precise. And I believe that the way in which we do this depends on our views of what cognition consists of. In this section I will sketch two different ways of unpacking the idea of cognitive extension. One is tailor-made for the functionalist view of cognition that underlies Clark-style artifact-extended cognition. The other is more suitable for Gallagher-style enactivist views of cognition—even though I am less sure he would accept it.

The meaning of cognitive extension that fits a functionalist outlook on cognition such as Clark's best is what I will label "implementation extension." According to functionalists, cognitive states and processes are to be characterized as functional role states and transitions from one set of functional states to another [this formulate is an attempt to cover as many variants of functionalism as possible, but at any rate machine functionalism (Putnam, 1967), psycho-functionalism (Fodor, 1968), and analytical functionalism (Lewis, 1972)].

[2]I am using scare quotes because I am not implying any commitment to mechanistic explanation in cognitive science. I will argue in this section that an enactivist view on cognition yields a different notion of cognitive extension than a functionalist view. My formulation must therefore be enactivist-friendly. Although I do not believe that mechanistic explanation and enactivism are enemies (Abramova and Slors, 2019), many enactivists do not accept a mechanistic style of explanation in cognitive science. By "mechanism" I mean something like processes that are responsible for the way a cognitive system functions.

That is, a mental process such as remembering is to be characterized in terms of the function it fulfils for an organism: storage and retrieval of information for the purpose of action control. Functional role states and functional processes are implemented or realized by physical structures that play the appropriate causal roles. Usually these are brain states and processes. The basic idea behind functionalism is that functional role states and processes are multiply realizable: the same function can be physically realized in different ways. And it is precisely this multiple realizability that is put to use in cases such as Otto and Inga, where the same functional process has two different realizations or implementations, one involving brain processes only, the other involving an item in the external world as well. Implementation extension is the idea that the realization or implementation base of functional role states and processes that are characteristic of human cognition includes items outside the brains and bodies of persons.

The notion of implementation extension is probably the most straightforward interpretation of extended cognition, so I will be brief about it. Two things are important to note. First, implementation extension fits really well with artifact extension, since physical artifacts are easy to imagine to be causally coupled with brains and bodies in ways that extend the implementation base of functional processes. It may fit with social extension as well, but as we have seen above this soon leads to an implementation base of extended cognitive processes that covers more than can possibly be said to belong to the cognition of a single person. Secondly, as second-wave extended cognition theories stress, extending the implementation base of functional processes may lead to new functional processes that have no mere brain-based parallel.

The second interpretation of extended cognition is less well-entrenched in philosophy of mind. It may be made compatible with a functionalist outlook, but is fits best with an enactivist notion of cognition. Briefly put, according to enactivists, cognition is a specific type of bodily engagement of an organism with the world [I will, again, try to formulate so as to cover most varieties of enactivism, including autopoietic (Thompson, 2007), sensory-motor (Noë, 2004), and radical (Hutto and Myin, 2013) enactivism]. Cognition is not a hidden layer between perception and action where the real thinking occurs. It is responding to the action-opportunities offered by the environment to an organism in such a way that the organism benefits, e.g., by sustaining its own organization. Cognition is a process that encompasses perception, action and bits of the world. A cognitive process is a specific type of interaction between an organism and the world. Extending a cognitive process in this sense is not extending a realization base of a functional role (because there is no such thing according to enactivists), it is extending the part of the world we can engage with. Differently put, it is increasing the *impact* that a cognitive engagement with the world has, for example on the further action possibilities offered by the environment to the acting organism. Extending the impact of engagements can be achieved by involving specific artifacts in the interaction, but it can also—crucially—be achieved by embedding the interaction in specific social practices.

Some examples of the way in which social practices or institutions extend the impact of cognitive engagements with the world may help to get the idea across. The example of fencing off a piece of land is a good case in point. This relatively simple engagement with the world has the much wider impact of avoiding trespassers on your land only because it is embedded in a context of cultural conventions. But here the impact-extension is still relatively modest. Compare, for example, the process of signing a contract. This is, again, a relatively simple action. But given the legal system in which it is embedded—a system of rules and a practice of using and reinforcing them—as a cognitive engagement it can have a very wide impact. It will change the rights and obligations of the signers, making them house-owners, companions in a firm, employees, etcetera. Or think of a voting process in the board of a large company on a possible reorganization. With five votes for and five against, your vote is the last. By simply raising a hand, you set in motion a large reorganization. Raising a hand is a very modest engagement with the world. But by embedding it in complex of social practices—cultural conventions, economic processes, and legal transactions—its impact is massively extended[3].

Implementation extension and impact extension are very different forms of cognitive extension—or so I will argue. Many cases of implementation extension start with a pre-existing brain-based cognitive process that is extended by adding external items to the implementation base of these processes. Otto and his notebook are the perfect case in point. Impact extension, specifically if this is social extension, by contrast, involves the creation of new cognitive processes that match pre-existing social practices. As Gallagher rightly stresses, socially extended processes such as signing contracts or voting are not even intelligible in abstraction from the social practices they are part of (see section "Cognitive Symbiosis, Weak and Strong"). Raising a hand or making a scribble on a piece of paper are not cognitive processes at all in abstraction from the relevant social practices that make these engagements instances of voting and signing a contract.

The fact that in "socially impact extended cognition" social practices precede the development of cognitive abilities that help individuals use *and* contribute to these practices suggests a completely different sense in which items outside our brains and bodies can be said to co-constitute our cognitive processes. This is not the type of constitution that is characteristic of the functionalist outlook, where constitution is explained in terms of realization or implementation. Rather than saying that a cognitive process—characterized in functional terms—is constituted by the physical structures that have the relevant causal-functional characteristics, the point here is that certain engagements with the world are parts of the collective behavioral patterns that instantiate a specific social practice. The context of such a practice is needed for these engagements to make them into what they are; to make raising a hand voting and scribbling on a piece

[3]One could say that these social practices/institutions allow a person to engage with a much bigger portion of the world. Thus, instead of impact extension we could also speak of "engagement extension." For the sake of simplicity I will use one label—impact-extension—only.

of paper signing a contract. In fact, it is more natural to say that these engagements contribute to the perpetuation of institutional practices than it is to say that these practices extend these engagements—though it ultimately boils down to the same claim. It is exactly the fact that these engagements are contributions to institutional practices that explains their (hugely) extended impact, and this makes institutions *co-constitute* these engagements *as* the cognitive processes they are—voting and signing a contract.

I believe that this explanation of what it means to say that social institutions co-constitute some of our cognitive processes is more informative and better applicable to the idea of socially extended cognition than the definition of constitution referred to by Gallagher, 2013 himself in a footnote (2013, 6). According to that definition, "P is a constitutive element [of X] if P is part of the processes that produces X" (De Jaegher et al., 2010, 443). The problem with this definition is that it implies that "[t]he set of all the constitutive elements is the phenomenon itself" (De Jaegher et al., 2010, 443). The suggested identity relation is problematic, since identity is symmetric. But when an act of signing a contract is co-constituted by a legal system, "the set of constitutive elements" is vastly more encompassing than "the phenomenon itself." A definition of constitution in terms of parts that jointly make up a phenomenon fits better with implementation extension than it does with impact extension. In fact, impact extension involves a notion of constitution that employs the inverse relation: a social institution co-constitutes an engagement with the world as a given cognitive process (by massively extending its impact) not because the institution is part of the engagement, but because the engagement is part of the institution.

If the claim of socially extended cognition is understood in terms of impact extension, the problem of cognitive bloat does not arise. For the claim is no longer that a given social institution is part of a cognitive process, but rather that a cognitive process is part of a social institution.

CAUSALITY, COORDINATION, AND RECIPROCAL COGNITIVE DEPENDENCY

The notions of causal coupling and functional integration are perfectly at home in the context of implementation extension. The implementation base of a given cognitive process, understood along functionalist lines, consists of causally connected parts that together realize a given functional state or process. Such a base can be extended by causally coupling with further items so that its functionality is increased. This is functional integration. But what about impact extension? As discussed in section "Introduction," the notion of coupling is used in the context of socially extended cognition as well. If socially extended cognition is an instance of impact extension, this would suggest that impact extension hinges on causal coupling as well. Although I will not deny that impact extension involves causal coupling, my claim is that causal coupling is not the most important principle behind impact extension.

The most important principle behind the cognitive extension offered by institutions—impact extension—is the normativity that comes with the organization and coordination of tasks, roles and actions that is characteristic of an institution.[4] What extends the impact of putting a scribble on a piece of paper so that it makes me the owner of a house, say, is not just the causal contact of the pen on the paper, nor even the causal contact between the paper and the brain of a notary, a solicitor, a broker, a former owner, or a potential squatter, but the fact that the paper is treated by these as conferring specific rights and obligations that are respected by all. This is a normative practice—a practice in which keeping to specific organized roles is the norm and in which deviation is sanctioned. A legal system is first and foremost a collectively enacted system of coordinated actions. And this coordination is the result of the perceived normativity of the rules governing the system. This abstract description applies to all social institutions that can be said to extend our cognitive abilities. The main differences between legal systems, educational systems, systems of cultural conventions and other "mental institutions," are in the rules that govern the different systems, connected with the goals of the systems, and in the ways in which deviation from norms is sanctioned (Bicchieri, 2005).

The generally perceived normativity of the rules that govern a given institutional practice—whether enforced or not—allows for the kind of predictability of a given practice that is a precondition for the idea of socially extended cognition. The predictability of the proceedings of a given social institutions is the equivalent of the reliable availability and automatic endorsement of notes in Otto's notebook. Without sufficiently felt normative force of the principles governing an institutional practice, a practice ceases to be reliable enough to extend the impact of cognitive engagements with the world. If only some people respect the rights conferred to me on the basis of a signed contract, a fading social institution will no longer extend my cognitive engagements and signing a contract will no longer make me a house owner.

The emphasis on normativity, organization and coordination is intended to contrast with the mere mechanical causality that governs artifact-extended cognition. The structures that socially extend our cognitive abilities consist not just of physical artifacts but of (many) other people. We can be causally coupled with other people in many ways, but unless these other people behave in more or less predictable ways, such coupling will not yield cognitive extension. For this we need rules or organizing principles with normative force. In an earlier paper (Slors, 2019). I tried to capture the importance of normativity, organization and coordination and to contrast it with mechanical causality by using a distinction between functional integration (or causal coupling) and what I labeled "task-dependency," the fact that socially extended cognitive engagements with the world only make sense in the context of a social institution. I argued that

[4]There are other kinds of normativity. In particular, there are norms for the manipulation of cognitive devices that I would count as instances of artifact-extended cognition. Menary (2010, 238–241) gives an instructive overview of these. As Menary himself emphasizes, however, such cognitive normativity should be distinguished from social normativity. The contrast I wish to make between causal and normative connections pertains to social normativity.

socially extended cognition is less characterized by functional integration and more by task-dependency. Gallagher et al. (2019) accept the distinction between functional integration and task-dependency, but are critical of the claim that socially extended cognition is characterized by low functional integration and high task dependency. They argue that:

> An attorney, for example, has to make the system work by doing certain things that require material engagement with papers, law books, courtrooms, and many other people. What she does may be defined in terms of specific tasks, but those tasks are accomplished only by engaging with instruments and people, and often in flexible and creative ways. Contracts and written (official) documents are instrumentally functional and, at the same time, they are "pieces" of the legal structure that in some cases predefine or scaffold the roles of individuals. That is, at the same time, they are, from the individual's perspective, functionally instrumental for extending legal reasoning and, from the systems perspective, constitutive parts of the legal structure (Gallagher et al., 2019, 8).

I believe they are right. The contrast between artifact-extended and socially extended cognition that is the topic of this paper need not hinge on claims about low functional integration in socially extended forms of cognition. The point should simply be that even though material engagement is crucially important (Malafouris, 2013), social engagement is different from mere causal coupling, because it involves other minds, organization, coordination and normativity. In fact, this normativity carries over to the material engagements that (Gallagher et al., 2019) are correct to claim are important parts of social institutions as well. An attorney's engagement with a law book is subtly but crucially different from Otto's engagement with his notebook due to its being used in the context of reinforcing norms, rather than merely manipulating information. The normative dimension of the practice of law enters into the attorney's engagement with the law book, and this is absent in Otto's interactions with his own notebook.

So, my claim is that socially extended cognition differs from artifact extended cognition because the extending structures in the case of socially extended cognition contain (many) other minds, the required predictability of which can only be due to shared rules and principles that define a given social institution, which are perceived to have normative force. Socially extended cognition adds normativity to the causal coupling with other people and with artifacts that socially extended cognition shares with artifact-extended cognition.

There is another difference between socially extended cognition and artifact-extended cognition that is implied by the above discussion, but not made explicit. Artifact-extended cognition is asymmetrical or non-reciprocal. Otto's mind is extended by his notebook, not the other way around. The material structuring of actors in 16th century London allowed them to memorize more than ten Shakespeare plays simultaneously and thus extended their minds, but not the other way around. By contrast, socially extended cognition is reciprocal. Social institutions extend our cognitive abilities because we contribute to the practices that define these institutions. By contributing we co-constitute these institutions just like these institutions co-constitute our cognitive abilities (see previous section "Implementation-Extension and Impact-Extension"). And since the cognitive abilities of others are just as well co-constituted by social institutions as ours, we contribute to the cognitive extension of others just as they contribute to ours. Social extension of cognition is reciprocal co-constitution of cognitive abilities.

Given that socially extended cognition is different from artifact-extended cognition—it involves an important normative component, and it is characterized by impact-extension rather than implementation extension, which is reciprocal rather than unidirectional—it may be useful to give it a label of its own. Calling both type of cognition "extended" glosses over important differences. Given the reciprocal cognitive dependency in socially extended cognition, I believe the term "symbiotic cognition" is apt.

COGNITIVE SYMBIOSIS, WEAK AND STRONG

Let me summarize the defining features of symbiotic cognition that follow from the above discussion. I will first define what I will label "weak symbiotic cognition," in abstract terms, briefly comment on the defining features and discuss an example as illustration. I will then argue that it is possible that there are forms of socially extended cognition that do not meet the requirements for symbiotic cognition. Weak symbiotic cognition does not hinge on social institutions. The kind of socially extended cognition referred to by Gallagher, by contrast, exemplified by the examples of signing a contract and voting by raising a hand above, does involve social institutions. This is what I will label "strong" or full-blown symbiotic cognition. It involves a further defining feature that I will discuss and elaborate on at the end of this section.

Weak symbiotic cognition, as I will use the term, is:

(i) a form of *socially* extended cognition,
(ii) that involves *impact extension* rather than implementation extension,
(iii) that involves *normativity* in the interactions between persons on top of causal coupling,
(iv) that involves the *reciprocal* co-constitution of cognitive abilities between persons,
(v) where the social co-*constitution* of cognitive abilities is due to the fact that cognitive processes are shaped as parts of *pre-existing social structures*.

Features (ii–v) are further specifications of (i). As I will argue below, it is defensible to claim that some forms of cognition are socially extended without satisfying (ii–v). Features (ii–v) are strongly interconnected; they highlight different aspects of weak symbiotic cognition, but seem to be a package deal, rather than separate individual necessary conditions.

Feature (ii) has been discussed above. It is important to note that impact extension requires a pre-existing social structure. Without a pre-existing legal system, for example, putting a

scribble under a document would not amount to signing a contract and becoming a property-owner.

Feature (iii) follows from the distinction between impact extension and implementation extension discussed above. Initiating cognitive engagements because of their assumed extended impact (say, raising a hand in a vote) anticipates predictable behavior of others in the same social structure. This predictability hinges on the felt normativity of structure-sustaining behavior [see (v)].

Feature (iv) does not imply that reciprocal co-constitution of cognitive abilities is necessarily symmetrical. It may well be that by playing different roles in the same social structure we co-constitute different cognitive abilities in each other.

Feature (v) is deliberately vague about the nature of social structures. The term might refer to social institutions, but this need not be the case. There is structure in human interactions when there are identifiable roles that interact in ways that allow us to discern regularities. The sense in which social structures "pre-exist" before symbiotic cognitive processes can occur is metaphysical, and not necessarily temporal (though in most instances it will be temporal as well): without the context of a social structure, a symbiotic cognitive process cannot exist as such.

Various forms of collective cognitive activity satisfy (i–v), without being instances of the type of cognition Gallagher refers to, i.e., cognition in the context of social institutions. Group-memory is a well-researched case in point. While some researchers argue that memory storage and retrieval by groups is impaired relative to the sum memory abilities of the individual members of a group (Pavitt, 2003), there is considerable research that shows that group-level performance adds to the sum of individual performances (see Theiner et al., 2010, 388–389 for a brief but well-argued overview; see Theiner, 2013 and Arango-Muñoz and Michaelian, 2020 for detailed analyses). Daniel Wegner has probably provided the most famous example of this with his notion of a "transactional memory system," consisting of two or more individuals who have acquired specific, often implicit routines that allow them to divide and combine cognitive labor efficiently. Thus, long-term married couples are capable of remembering much more together than separately (Wegner, 1986). It is important, in such cases, that we do not disrupt the ingrained routines. Assigning a different, new division of cognitive labor, for example, reduces the collective memory capacity of couples demonstrably (Wegner et al., 1991). These routines are instances of the pre-existing social structures referred to in (v).

In general, task division in couples that live together for some time often rigidifies into shared routines, that are usually based on tacit knowledge of individual proclivities and talents, and that usually amount to the automatic complementing of each other's cognitive efforts. Such routines would make the couple into a symbiotic cognitive systems in terms of the above definition. Let me take the following, simplified case as an example: when on vacation, my wife always takes care of train- plane- or boat-tickets and the planning of when we should go where and what to see, whereas I do navigation and hotel arrangements, tents (in which case my wife determines the campsite) and guesthouses.

This (simplified) arrangement satisfies (ii–v):

(ii) My actions of arranging tent-gear and navigating result in having a complete vacation, including interesting trips, a nice campsite, a boat trip, etcetera, because they are done in the context of a (weakly) symbiotic system. This is a form of impact extension; outside of this context the same actions would not have that effect.

(iii) There is most certainly a kind of normativity involved in our division of cognitive labor. This is based on precedent and on shared assessment of talents which leads to mutual expectations.

(iv) We co-constitute each other's cognitive abilities. By dividing complementary cognitive tasks and by using many automatized interaction routines that let us share information when necessary (and not when not necessary), we co-constitute each other's ability to realize a full vacation with roughly half the effort.

(v) These routines—our implicit knowledge of the way in which we divide cognitive labor and share results when necessary—counts as social structure of the relevant kind (i.e., supporting reciprocal co-constitution of cognitive abilities).

Are there forms of socially extended cognition that do not satisfy (ii–v)? I believe that that is possible, depending on how widely we apply the term "socially extended." For example, the relation between a student and a teacher might be described as socially extended cognition—the student's cognition is extended by the teacher's (note that nothing in this paper hinges on calling this an instance of extended cognition). Likewise, a reader's cognitive abilities might be thought of as being extended by the cognitive activities of a writer. There are reasons to be cautious here in describing such cases as instances of socially extended cognition,[5] but even if we disregard these, such cases are not instances of symbiotic cognition. For first, and most importantly, these relations do not satisfy (iv): the cognitive extension is a one-way affair and not reciprocal—teachers extend the cognition of students, but not vice versa and writers extend the cognitive abilities of readers, but not vice versa. This might be argued to affect (ii), (iii), and (v) as well. To start with (v), the social structures involved are not structures of the right kind because they do not involve mutual dependency. Also, these relations do not involve the right kind of normativity. There may certainly be normativity involved in these relations or in playing the relevant roles involved, but not necessarily normativity of the kind that renders the behavior of others predictable so that cognitive engagements by the agent are impact-extended. Which means that (ii) is not satisfied either. Having said that, though, nothing hinges on these assessments of the applicability of (ii), (iii), and (v); the non-applicability of (iv) suffices to rule out these cases as cases of symbiotic cognition.[6]

[5]One problem here is that while there is no impact-extension involved, it would be somewhat odd to speak of implementation-intention, unless we want to include other people in the implementation base of one' own mental processes.

[6]The point that is made here about socially extended cognition can also be made about affective social scaffolding. Stephan and Walter (2020, section 4), mention examples such as seeing a psychotherapist, confessing to a priest, the emotion

I have labeled forms of symbiotic cognition that do not involve social institutions "weak symbiotic cognition," because they differ in one important respect from socially extended cognition of the type Gallagher discusses. I believe that the discussion of the previous sections suffices to show that (i–v) apply to Gallagher's cases. But these cases have a striking feature that is lacking in the case of a married couple jointly planning and having a vacation or the case of collective memory. The cognitive engagements Gallagher discusses are only intelligible *within* the context of their respective institutions. Many of our daily cognitive activities have this property. Signing a contract is not intelligible in abstraction from a legal system, voting is not intelligible in abstraction from a social structure which allows for joint decision making, being polite by shaking hands is not intelligible in abstraction from a system of cultural conventions, etcetera. What I will label "strong" or full-fledged symbiotic cognition, then, adds one more requirement to (i–v):

(vi) Cognitive processes are *possible* and *intelligible* only within the context of a social institution.

Crucially, the example of married couples with ingrained automatized routines, or transactional memory systems, are not examples of strong symbiotic cognition. For the individual cognitive processes within such symbiotic systems *are* intelligible in abstraction from the system. My activity of navigating or booking a hotel does not require my wife's activity of planning trips and booking tickets to be intelligible. Neither does the individual memory-contribution of an individual to a transactive memory system require reference to other people to be intelligible as a memory process.[7] Weak symbiotic cognitive systems combine individual cognitive processes, that do not require the system to exist, into a larger system that is beneficial to participants. Strong symbiotic cognition, by contrast, cannot be reduced to a collection of individually intelligible cognitive processes. It is only in connection with the whole system that strongly symbiotic cognitive processes are cognitive processes at all. It is not just that the whole is more than the sum of its parts (see footnote 7), the point is rather that there are no identifiable relevant parts without the notion of the whole.

Take the case of signing a contract again. What it *means* to sign a contract involves reference to a very complex social structure in which rights and obligations exist and can be changed. "Rights and obligations" refers to very specific norm-guided, socially structured behavior. It is not possible to identify that behavior fully, in turn, without referring back to contracts. The roles and regularities of the social structures involved in strong symbiotic cognition are *holistically* inter-defined (Slors, 2019). To define the role of a barrister, one has to refer to the rule of law, and to roles of citizens, judges, clerks and many others. And to define these other roles, reference to the roles of barristers will have to be made. To define the role of a board member, one has to refer to the whole organizational structure of a company.

For the type of roles and regularities to exist that can and need to be holistically inter-defined, a certain degree of complexity is required. Strongly symbiotic systems, then, are likely to be much larger systems than transactional memory systems. A legal system, typically, is enacted by a whole society. A company is enacted by a very large group of people, and can exist only within an economic arrangement that involves whole countries. Strong symbiotic cognition, then, is not just a more stringent sub-variety of weak symbiotic cognition.

The holistic inter-defining of roles and regularities implies that strongly symbiotic cognitive engagements or processes are necessarily aimed at accomplishing a given state of affairs *within* the relevant symbiotic system. Any cognitive engagement that counts as executing a system-defined role implies the involvement of other people playing their respective roles in that same system. Signing a contract is what it is because it affects the roles, obligations and rights of other people (a former house owner, say, can no longer determine what is to be done with a house once it is yours, due to you signing a contract; she can no longer determine this *as* a citizen who falls within the same legal system as you do). Shaking hands as a greeting opens up a new space of social interaction possibilities due to the fact that those involved all participate in the same system of cultural conventions—it is a "move" *within* the "game" of social etiquette that is meaningless or weird to anyone who does not share your conventions.

Feature (vi), then, transforms weak symbiotic cognition into a qualitatively different kind of cognition. If (vi) is added to (ii–v), and the five features together are taken as interconnected, then (ii–v) are substantially strengthened. Of course feature (v) is further defined by limiting the pre-existing social structures to social institutions. But this affects the other features too. Impact extension (ii) within a strongly symbiotic system is substantially more encompassing than impact extension in a weakly symbiotic system. Setting a whole reorganization of a company in motion by raising a single hand illustrates the point. This is a different scale of impact-extension than having a whole vacation with half the work. (iii) The normativity involved in social institutions is not merely dependent on precedent and implicit assessment of talents and proclivities. Precisely because it applies to much larger groups, it is usually reinforced, either explicitly, as in legal systems, or implicitly, as in a system of social etiquette. (iv) The co-constitution of cognitive abilities in strongly symbiotic systems is much more elaborate than in weakly symbiotic systems. First of all this is because many more people are involved. But secondly this is because most social institutions, instill a wide range of "new" cognitive abilities in those who help to enact them.

As said, I take Gallagher to refer to strong of full-fledged symbiotic cognition in his discussion of socially extended

regulation involved in infant-caregiver interactions (see also Krueger, 2013), and the transformative effect of social media on our affective mindset (or our mindset in general). However, in all but the last of these examples, the reciprocity that is characteristic of symbiotic cognition is absent or so much diminished that I would consider them borderline cases at best.

[7]This is not to say that collective memory cannot be an emergent process. It can. Emergence hinges on the way that an overall process such as collective remembering depends, ontologically, on its constituent processes (see Arango-Muñoz and Michaelian, forthcoming, sections 11.3.2 and 11.4 for a discussion of different forms of emergence in the context of collective memory). It does not require that the constituent processes be definable or intelligible only in ways that refer to the overall process whose emergence they contribute to.

cognition. In the remainder of this paper I will refer to this type of cognition simply as "symbiotic cognition."

SYMBIOTIC COGNITION, COGNITIVE INTEGRATION AND DISTRIBUTED COGNITION

So far, I have limited the discussion to the literature on extended cognition, arguing that symbiotic cognition differs from "normal," artifact-extended cognition in some important respects. There are other theories about the essential embeddedness of our cognitive systems. Richard Menary's notion of cognitive integration does emphasize the expansion of our cognitive repertoire by engaging with a wide variety of cultural items, including social structures, but without making claims about the extension of our cognitive systems as such. Edwin Hutchins' notion of socially distributed cognition, by contrast, allows for whole social institutions to count as cognitive systems. I have argued that the literature on extended cognition has swept an important distinction under the carpet; it has not sufficiently recognized that socially extended cognition is—at least very often—a type of cognition of its own, fundamentally different from artifact-extended cognition. But it may well be that this distinction is respected by the notions of integrated cognition or distributed cognition. I which case I may have said nothing new. I will briefly argue, however, that neither cognitive integration, nor distributed cognition is very sensitive to the distinction I have argued for above.

The idea of cognitive integration is in many respects very close to the idea of extended cognition. Cognitive integration is also close to the enactivist view in that it emphasizes that cognition consists of bodily manipulations of the world, often involving man-made cognitive devices (alternatively, it may, according to Menary, also consist of mental simulations of such manipulations). Cognitive processes are cognitive practices, and these can be hugely expanded by involving a host of different items. The items mentioned in the cognitive integration literature fall in the same (wide) range as the devices referred to by extended cognition theorists. The crucial difference with extended cognition is that while according to Menary items such as linguistic symbols, smart phones, abacuses and social institutions allow for a whole new range of cognitive practices, they are enabling conditions for such practices, rather parts of our minds. In this respect, Menary is closer to those who argue that external devices scaffold our cognition, rather than extend it (e.g., Sterelny, 2010).

It should be noted that the notion of cognitive extension that Menary rejects is a variant of what I have labeled "implementation extension" above. Even though he tends toward an enactivist notion of cognition rather than a classical functionalist one, he still speaks of cognition "supervening" on a realization base and thinks of cognitive extension in terms of enlarging this base. This raises the question whether perhaps impact-extension might be compatible with the idea of cognitive integration. The similarity between the enactivist notion of cognitive engagement and Menary's notion of cognitive practices might suggest this. Indeed, there are clear similarities. Menary speaks of the "transformation" of our minds by cognitive artifacts and our interactions with them in a way that suggests that manipulating these artifacts has a cognitive yield in the context of cognitive practices that the same manipulation would not have outside of such practices. The impact of an ignorant infant who happens to manipulate numeric symbols such that they accidentally represent a calculation differs from the impact of a mathematically trained person who performs the same manipulation. This is akin to the difference between someone coincidentally putting a scribble on a piece of paper and someone signing a real contract. The practice extends the impact of the manipulation.

However, even though it may be argued that this type of "transformation" of cognitive processes is very much like impact-extension, this does not mean that the idea of cognitive integration already contains or implies the notion of symbiotic cognition. On Menary's view, all cognitive integration is somewhat like impact-extension. The contrast between socially extended/integrated and artifact-extended/integrated cognition—or between what I would prefer to call extended and symbiotic cognition—is not made. Hence, in this respect it will not help to abandon extended-cognition talk in favor of cognitive integration.

What about socially distributed cognition? On Hutchins' original proposal, (Hutchins, 1995) socially distributed cognition is a view on cognition that is much like the idea of group minds (Theiner et al., 2010). The point of this view is that it is perfectly possible for a group of people to jointly carry out certain cognitive tasks. Can social institutions be viewed as cognitive systems? On the wide characterization of "cognitive system" employed by Hutchins, 2014 in his later work (e.g., 2014), they can. For here the criterion is not that a system has a given task (as in Hutchins earlier work), but that it consists of integrated cognitive elements such that i.e., multiple human beings in conjunction with a cultural niche replete with cognitive artifacts counts as such a system. Hutchins speaks of "a cognitive ecosystem." A social institution can certainly be viewed as a cognitive ecosystem. Cognition in a cognitive ecosystem is not implementation-extended, but impact-extended. Like symbiotic cognition, and unlike extended cognition, Hutchins emphasizes that distributed cognitive systems have no center—there is no one brain that is extended by others, but there is what I called reciprocal extension.

In many respects, therefore, symbiotic cognition can be viewed as a variant of the cognitive ecosystems view implied by later versions of the idea of distributed cognition. The one thing that is missing, however, like in the case of cognitive integration, is the relevant contrast between extended and symbiotic cognition. Hutchins (2014, 36–38) still thinks of extended cognition as a possible variant of distributed cognition. Thus, he ignores the difference between causal coupling and reciprocal social-normative coupling that involves organization and action coordination. To sum up, then: some elements of symbiotic cognition can be found in the ideas of integrated and distributed cognition, but the relevant contrast between

symbiotic and extended cognition that I have been arguing for in this paper is still absent.

CONCLUSION

I have argued that there is an important distinction between cognitive extension as the extension of the causal-functional implementation base of cognitive processes, which is best applicable in cases where cognition is extended by physical artifacts only, and cognitive extension as the idea that our cognitive engagements with the world have massively enhanced impact in the context of normative, rule-based coordination of actions in a social practice. Though both types of cognition might equally well be called "extended," they are extended in radically different ways. In order to mark this difference, and given the reciprocal cognitive co-constitution between humans in impact-extended cognition, I have proposed to label what is now known as socially extended cognition "symbiotic cognition."

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I would like to thank the two referees for this journal for their helpful comments and questions. Thanks for pushing me to write section "Cognitive Symbiosis, Weak and Strong."

REFERENCES

Abramova, E., and Slors, M. (2019). Mechanistic explanations for enactive sociality. *Phenomenol. Cogn. Sci.* 18, 401–424. doi: 10.1007/s11097-018-9577-8

Adams, F., and Aizawa, K. (2001). The bounds of cognition. *Philos. Psychol.* 14, 43–64.

Adams, F., and Aizawa, K. (2008). *The Bounds of Cognition*. Malden MA: Blackwell.

Adams, F., and Aizawa, K. (2010). "Defending the bounds of cognition," in *The Extended Mind*, ed. R. Menary (Cambridge, MA: MIT Press), 67–80. doi: 10.7551/mitpress/9780262014038.003.0004

Arango-Muñoz, S., and Michaelian, K. (2020). "From collective memory to collective metamemory?," in *Minimal Cooperation And Shared Agency*, ed. A. Fiebich (Berlin: Springer).

Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York, NY: Oxford University Press.

Clark, A. (2010). "Memento's revenge: the extended mind extended," in *The Extended Mind*, ed. R. Menary (Cambridge, MA: MIT Press), 43–66. doi: 10.7551/mitpress/9780262014038.003.0003

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

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009

Dennett, D. C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge MA: MIT Press.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge MA: MIT Press.

Fodor, J. (1968). The appeal to tacit knowledge in psychological explanation. *J. Philos.* 65, 627–640.

Fuchs, T., and De Jaegher, H. (2009). Enactive intersubjectivity: participatory sense-making and mutual incorporation. *Phenomenol. Cogn. Sci.* 8, 465–486. doi: 10.1007/s11097-009-9136-4

Gallagher, S. (2013). The socially extended mind. *Cogn. Syst. Res.* 2, 4–12. doi: 10.1016/j.cogsys.2013.03.008

Gallagher, S., and Crisafi, A. (2009). Mental institutions. *Topoi* 28, 45–51.

Gallagher, S., Mastrogiorgio, A., and Petracca, E. (2019). Economic reasoning and interaction in socially extended market institutions. *Front. Psychol.* 10:1856. doi: 10.3389/fpsyg.2019.01856

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenol. Cogn. Sci.* 14, 577–598. doi: 10.1007/s11097-014-9355-1

Huebner, B. (2013). Socially embedded cognition. *Cogn. Syst. Res.* 2, 13–18. doi: 10.1016/j.cogsys.2013.03.006

Hurley, S. (2010). "Varieties of externalism," in *The Extended Mind*, ed. R. Menary (Cambridge MA: MIT Press), 101–153.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge MA: MIT Press.

Hutchins, E. (2014). The cultural ecosystem of human cognition. *Philos. Psychol.* 27, 34–49. doi: 10.1080/09515089.2013.830548

Hutto, D. D., and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds without Content*. Cambridge MA: MIT Press.

Hutto, D. D., and Myin, E. (2017). *Evolving Enactivism: Basic Minds meet Content*. Cambridge MA: MIT Press.

Krueger, J. (2013). Ontogenesis of the socially extended mind. *Cogn. Syst. Res.* 2, 40–46. doi: 10.1016/j.cogsys.2013.03.001

Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Austr. J. Philos.* 50, 249–258. doi: 10.1080/00048407212341301

Malafouris, L. (2013). *How Things Shape the Mind. A Theory of Material Engagement*. Cambridge MA: MIT Press.

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Basingstoke: Palgrave Macmillan.

Menary, R. (2010). "Cognitive integration and the extended mind," in *The Extended Mind*, ed. R. Menary (Cambridge MA: MIT Press), 227–244.

Menary, R. (2013). Cognitive integration, enculturated cognition and the socially extended mind. *Cogn. Syst. Res.* 25–26, 26–34. doi: 10.1016/j.cogsys.2013.05.002

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

Pavitt, C. (2003). Colloquy: do interacting groups perform better than aggregates of individuals? Why we have to be reductionists about group memory. *Hum. Commun. Res.* 29, 592–599. doi: 10.1111/j.1468-2958.2003.tb00857.x

Putnam, H. (1967). "The nature of mental states," in *Art, Mind, and Religion*, eds W. H. Capitan, and D. D. Merrill (Pittsburgh: Pittsburgh University Press), 12–23.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *J. Philos.* 101, 389–428. doi: 10.5840/jphil2004101826

Slors, M. (2019). Symbiotic cognition as an alternative for socially extended cognition. *Philos. Psychol.* 32, 1179–1203. doi: 10.1080/09515089.2019.1679591

Stephan, A., and Walter, S. (2020). "Situated affectivity," in *The Routledge Handbook of Phenomenology of Emotions*, eds T. Szanto, and H. Landweer (London: Routledge).

Sterelny, K. (2010). Minds: extended or scaffolded. *Philos. Cogn. Sci.* 9, 465–481.

Sutton, J. (2010). "Exograms and interdisciplinarity: history, the extended mind, and the civilizing process," in *The Extended Mind*, ed. R. Menary (Cambridge MA: MIT Press), 189–226.

Theiner, G. (2013). Transactive memory systems. A mechanistic analysis of emergent group memory. *Rev. Philos. Psychol.* 4, 65–89. doi: 10.1007/s13164-012-0128-x

Theiner, G., Allen, C., and Goldstone, R. L. (2010). Recognizing group cognition. *Cogn. Syst. Res.* 11, 378–395. doi: 10.1016/j.cogsys.2010.07.002

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology and the Sciences of the Mind*. Cambridge MA: Harvard University Press.

Tribble, E. (2005). Distributing cognition in the globe. *Shakespeare Q.* 56, 135–155. doi: 10.1353/shq.2005.0065

Wegner, D. M. (1986). "Transactive memory: a contemporary analysis of the group mind," in *Theories of Group Behavior*, eds B. Mullen, and G. R. Goethals (New York: Springer), 185–208. doi: 10.1007/978-1-4612-4634-3_9

Wegner, D. M., Erber, R., and Raymond, P. (1991). Transactive memory in close relationships. *J. Pers. Soc. Psychol.* 61, 923–929. doi: 10.1037/0022-3514.61.6.923

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 26 June 2020
doi: 10.3389/fpsyg.2020.01278

Overcoming the Past-endorsement Criterion: Toward a Transparency-Based Mark of the Mental

Giulia Piredda and Michele Di Francesco*

Nets Center, Department of Humanities and Life Sciences, Scuola Universitaria Superiore IUSS Pavia, Pavia, Italy

Starting from the discussion on the original set of criteria advanced by Clark and Chalmers (1998) meant to avoid the overextension of the mind, or the so-called cognitive bloat, we will sketch our solution to the problem of criteria evaluation, by connecting it to the search for a mark of the mental. Our proposal is to argue for a "weak conscientialist" mark of the mental based on transparent access, which vindicates the role of consciousness in defining what is mental without, however, identifying the mental with the conscious. This renovated link between mind and consciousness, spelled out through the concept of transparency, further develops some of our previous work on the topic (Di Francesco, 2007; Di Francesco and Piredda, 2012) and is partially inspired by Horgan and Kriegel (2008).

Keywords: cognitive bloat, mark of the mental, consciousness, extended mind, transparency, past-endorsement criterion

OPEN ACCESS

Edited by:
Achim Stephan,
University of Osnabrück, Germany

Reviewed by:
Riccardo Manzotti,
Università IULM, Italy
Tom Roberts,
University of Exeter, United Kingdom

****Correspondence:***
Giulia Piredda
giulia.piredda@iusspavia.it

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *26 February 2020*
Accepted: *15 May 2020*
Published: *26 June 2020*

Citation:
Piredda G and Di Francesco M (2020) Overcoming the Past-endorsement Criterion: Toward a Transparency-Based Mark of the Mental. Front. Psychol. 11:1278. doi: 10.3389/fpsyg.2020.01278

INTRODUCTION

Mind-extenders are quite common in contemporary philosophy of mind, and consequently many arguments have been advanced to free our minds from the boundaries of skull and body. Yet, even the most confident mind-extender has to admit that it is necessary to avoid an overextension of the mental (the so-called cognitive bloat, Rowlands, 2009). In this paper, we address this problem in connection with the search for specific Criteria to Avoid the Overextension of the extended mind (let us call them CAOs) proposed by Clark and Chalmers (1998). More specifically, our starting point will be the fourth criterion, the so-called past-endorsement criterion: we think that, by introducing a direct reference to consciousness among the CAOs, this criterion raises important problems, whose solution involves an analysis of the connections between the subpersonal extended vehicles of cognition and the conscious mind of the (extended) subject, which in turn requires an answer to the "mark of the mental" problem.

In the first part of the paper, we will review the problem raised by the past-endorsement criterion since its first appearance in Clark and Chalmers (1998) and retrace its fortunes and misfortunes in the subsequent literature. We will conclude that, even if the solution to the overextension problem offered by the criterion is not satisfying, this portion of the debate is important, in that it suggests the opportunity to further investigate the role of consciousness in distinguishing mental from non-mental resources.

In the second part of the paper, we connect this debate to the search for a mark of the mental, sketching our own solution—which we define as "weak conscientialism," based on some our previous works (Di Francesco, 2007; Di Francesco and Piredda, 2012; Di Francesco et al., 2016; Di Francesco and Tomasetta, 2017) and partially inspired by Horgan and Kriegel (2008). We draw some conclusions in the last paragraph.

THE PAST-ENDORSEMENT CRITERION AS A SOLUTION TO THE OVEREXTENSION OF THE MIND

Following a by now well-established interpretation of the literature on the topic, it is possible to individuate at least three different "waves" in the development of the extended mind theory (Menary, 2010; Gallagher, 2018): the first—in the original version by Clark and Chalmers (1998)—is based on the parity principle[1]; the second—championed by Menary (2007, 2010) and Sutton (2010)—is built around the concepts of integration and complementarity; and the third—still in lively development—starts with enactivism and is connected to the model of the mind inspired by predictive processing framework (Hohwy, 2013; Clark, 2016; Kirchoff and Kiverstein, 2018).

Although time flows, and theories undergo adjustments, it is possible that some "recalcitrant" problems resist the flow of the different waves. In this paper, we start with one of these recalcitrant problems, one that appears already in the seminal paper by Clark and Chalmers and that in the following years—despite the vigorous development of the debate—never attracted much attention, with a few exceptions (e.g., Rupert, 2004; Gertler, 2007; Roberts, 2012).

In the final part of their article, Clark and Chalmers discuss the scope of the extended mind thesis just stated and its potential consequences. To individuate potentially crucial points, they spell out the features involved in the case of *extended belief* they presented the well-known case of Otto's extended belief stored in his precious notebook. In subsequent literature, these criteria have been dubbed the "glue and trust" criteria (cf. Clark, 2010b); despite their fame, though, there remain unanswered questions regarding both their validity and their role. Here is the first appearance of the criteria:

> First, the notebook is a constant in Otto's life—in cases where the information in the notebook would be relevant, he will rarely take action without consulting it. Second, the information in the notebook is directly available without difficulty. Third, upon retrieving information from the notebook he automatically endorses it. Fourth, the information in the notebook has been *consciously endorsed* at some point in the past, and indeed is there as a consequence of this endorsement. (1998, p. 17, our italics)

The first three criteria—constancy in use, direct availability, and automatic endorsement—appeal to structural or functional features and have been considered as fairly reasonable. Basically, they mimic the normal relation between the conscious mind and its internal subpersonal underpinnings (Di Francesco, 2007). When subpersonal processes give input to the conscious mind, they do it in a systematic and direct way and their content is mandatory. It is assumed by default as a datum, poised for verbal report, reasoning, and so on. In this sense, the first three criteria try to mirror, at the causal level, some important phenomenological properties of the personal mind.

Many critics of the extended mind have highlighted that the first three criteria seem too easily satisfied, such that they would be insufficient to block what has been considered an undesired and implausible proliferation of alleged extended beliefs (Rupert, 2004, p. 401 ff.). Without the "conscious endorsement" requirement, in fact, we should consider as extended beliefs any information coming from a constantly consulted and trusted source: say, for example, a service that provides phone numbers in an efficient and trusted way or some easily accessible web pages. But would it be cognitively plausible to claim that Otto, even before consulting the service, already has beliefs about the phone numbers or about the easily accessible web pages? It seems that posing a more restrictive criterion, that Otto has endorsed a particular content in the past and has thus entertained an occurrent belief about that content, is a good way to avoid the "cognitive bloat" (Rowlands, 2010, p. 93). In other words, if one endorses the extended view of the mind, and accepts that the information stored in Otto's notebook counts as beliefs, there is a risk of "overextending" the mind: why stop there? Why not also allow all the resources Otto frequently uses among his extended mental states? The idea is that there should be a way to restrict the application of the extension only to *plausible* cases of extended belief.

The overextension of the mind is surely blocked by the fourth criterion Clark and Chalmers put in place: the past-endorsement criterion. According to it, in order for a specific content to be considered one of Otto's mental states, Otto should have consciously endorsed this content in the past, and this content, say an address, is now stored in Otto's notebook because of this process of conscious past-endorsement.

Now, while this further criterion eliminates any risk of overextending the mind, one may ask at what cost it does so. As a matter of fact, several problems concerning the past-endorsement criterion have been pointed out. Just after having presented it, Clark and Chalmers themselves (1998, p. 17) recognized the problematic status of this criterion when they observed that non-extended beliefs may be acquired via non-conscious processes, and imposing the additional conscious endorsement criterion only to extended beliefs would be at least arbitrary. On a more general note, the role assigned to consciousness by this criterion does not seem in line with the spirit of the extended mind framework, which tries to undermine the privilege to internal processes, like consciousness. This point has been made explicitly by Rupert (2004):

> If an extended (or any) belief requires conscious endorsement in order to be a genuinely held belief, and conscious endorsement is

[1] "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 part of the cognitive process" (Clark and Chalmers, 1998, p. 8).

> ultimately an internal process [. . .], then the traditional subject is privileged in a deep sense, after all (p. 404).

There are several observations deriving from this situation, and some of them will lead us toward the next section, dedicated to the search for a mark of the mental as a solution to the problem of the overextension of the mind.

The first concerns the role of consciousness: even if it is true that we acquire and form beliefs also unconsciously, and it is implausible to assume two different generation processes for extended and non-extended states, it is possible to interpret the emergence of the topic of consciousness in the debate about the extended mind as non-arbitrary[2]. The idea is that the fourth criterion is perhaps too strong, and unacceptable, as it is, but we feel that the discussion it raises is important in that it suggests a role for consciousness in defining what is properly mental. We will return to this notion.

Second, this is not the only occasion in which Clark seems to have a prudential attitude and defend the priority of the "organism-centered," even if not "organism-bound," cognition (Clark, 2008, p. 123). In the further literature on extended mind, Clark and Chalmers have been criticized for this attitude, defined as "too Cartesian," which would be entailed—according to many—by the parity principle (cf. Sutton, 2010, in Gallagher, 2018, p. 430; Wheeler, 2010).

Third, another dimension that is missing in the first three criteria, while well represented by the fourth, is a historic dimension: that is, the fact that the agent has been acquainted with some contents and—partly because of this—we could attribute these contents to him. However, a historic solution is not the only available. Also, we will offer an alternative functional solution.

The curious thing is that, although the debate on the extended mind has flourished in the last few decades, a thorough discussion on the issue of criteria evaluation—and particularly the status of the fourth criterion—is still lacking[3]. As we have seen, the problematic status of this criterion was promptly acknowledged by Clark and Chalmers (p. 17), who, after warning the reader, left the criterion in a sort of "theoretical limbo." As noted before, a full-blown criticism of the criterion was later developed by Rupert (2004). In the further literature, the criterion was at times mentioned, at times it was missing (Menary, 2010, p. 424; see Clark, 2010b, p. 50; Gallagher, 2018), while in Clark (2008) the fourth criterion was treated as problematic, but nevertheless relevant: "the 'past conscious endorsement' criterion looks too strong. On the other hand, to drop this requirement opens the floodgates to [. . .] an unwelcome explosion of potential dispositional beliefs" (p. 96). However, as far as we know, the topic has never been fully elaborated by Clark and Chalmers in their subsequent works.

In this paper, we will get our chance to sketch a solution to this discussion, connecting the missing (or underestimated) debate on the fourth criterion to the fundamental issue of the mark of the mental. Our "sketch" will focus on the role of transparent access (Clark, 2004, 2008; Wheeler, 2019)—a fundamental feature of consciousness—in defining what is mental, thus contributing to the issue of the mark of the mental. In our view, the lack of analysis dedicated to the past-endorsement criterion is revealing of a missing analysis of the relation between the extended mind and the role of consciousness. We believe that, within the extended mind framework, the lack of a serious analysis of the role of consciousness in marking the mental opens the door to the risk of overextension and thus leaves the entire framework wanting. A proper treatment of these important points of connection is due.

THE MARK OF THE MENTAL

From the Criteria to Avoid Overextension (CAOs) to the Mark of the Mental

The connection between the CAOs and the mark of the mental is, in a sense, direct. The four CAOs were introduced by Clark and Chalmers to avoid mental overextension, and having a mark of the mental seems an immediate way to succeed in this goal. Imagine that we want to know if a certain state, event, or process is a mental item. If we had a mark of the mental, we would only have to check whether the item in question meets the criteria set by the mark.

Among the criticisms addressed to the first wave of extended mind, the lack of such a mark of the cognitive or mark of the mental has been one of the most significant (see Adams and Aizawa, 2001, 2008; Piredda, 2017 for discussion)[4]. The idea is

[2]Interestingly, Clark has explicitly defended internalism regarding consciousness (see Clark, 2009, 2012). The debate about extended consciousness is still open (e.g., Lycan, 2002; Vold, 2015; Kirchoff and Kiverstein, 2018; Chalmers, 2019; Manzotti, 2019), and the possibility of extending consciousness would bring completely different solutions to the problem solved by the past-endorsement criterion. Unfortunately, discussing these alternative possibilities would lead us astray from the topic of this article.

[3]Among the few exceptions is Gertler (2007). In her paper, she finds a way to block the overextension of the mind by blocking the extended mind itself, criticizing one premise of the argument for it. The result is an argument for a "narrow mind," according to which the mental is restricted to the conscious. While we find her point of view undoubtedly interesting, we do not agree with her conclusion—we would like to find a way to resist the undesired overextension of the mind, maintaining the existence of unconscious mental states and processes.

[4]In this paper, we shall use "mark of the mental" and "mark of the cognitive" as essentially synonymous expressions. The reason for this apparently objectionable choice is that there is no firmly established use of these two expressions in the debate on the extended mind. Generally speaking, "mental" has a broader meaning, and "cognitive" may refer to a subset of mental phenomena. Another difference (aligned, perhaps, with Clark and Chalmers' approach) is that "cognitive" may be reserved for subpersonal "intelligent" processing (as in the Tetris example) and "mental" for (potentially) conscious states (such as Otto's and Inga's beliefs). Most of the literature on the extended mind has focused more on the mark of the cognitive than on the mark of the mental. While the problem of clearly distinguishing between the two lies beyond the scope of this paper—and actually concerns most of the literature on the extended cognition/mind debate—we believe that the terminological choice between "cognition" and "mind" will depend, at least in part, on the philosophical taste and tradition of the author: a philosopher of cognitive science is more likely to talk about cognition, while it is more probable that an analytic philosopher, or a follower of the phenomenological tradition, or even a follower of radical enactivism, will talk about the mind and the mark of the mental. The distinction appears to be more sociological than metaphysical, so to speak. Nevertheless, we also believe that, from a substantial point of view, the debate on the "mark of the cognitive" developed in the literature on the extended mind, to which we refer in this section, is extremely relevant for the issue of the mark of the mental: the positions of Clark, Adams, and Aizawa

that "causal coupling" alone, even if constrained by the three "glue and trust" criteria, is not sufficient to individuate genuine examples of *cognitive* or *mental* activity: one would need a mark of the mental in order to discriminate the cases to be rightly counted as such.

While Adams and Aizawa (2001, 2008) provide a mark of the cognitive based on the notion of intrinsic content, Clark and Chalmers (1998) do not seem to provide such a mark. It is true that the problem of the mark was not mentioned in the 1998 paper, but Clark later acknowledged this problematic point and has dedicated some thoughts to it (Clark, 2008, 2010a,b,c). His ideas about the topic are oriented to a purely minimalist and functionalist position—an interesting approach, but one that is not able to solve on its own all the problems of overextension.

First of all, in Clark's view, what is cognitive or non-cognitive is not the single component of a certain process, but rather the process as a whole, which must be involved in supporting intelligent behavior:

> What makes a process cognitive [. . .] is that it supports intelligent behavior (Clark, 2010a, p. 92).

> The study of mind might [. . .] need to embrace a variety of different explanatory paradigms whose point of convergence lies in the production of intelligent behavior (Clark, 2008, p. 94)[5].

Thus, according to Clark, the processes could in principle be implemented by various kinds of substances (biological or artificial substrates, as well as external resources), because what defines something as cognitive or mental is neither the substance that realizes it nor the detailed causal dynamics that characterize its workings. In this sense, cognitive or mental processes in the extended framework are individuated on the basis of coarse or common-sense functional considerations concerning cognitive processes such as memory, understanding, categorization, reasoning, etc.

> It is the coarse or common-sense functional role that, on this model [...], displays what is essential to the mental state in question (Clark, 2008, p. 89).

The reference to the causal relationship as the starting point of the analysis on mental reality is, after all, at the base of the (extended) functionalist intuition:

> What makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body (Clark and Chalmers, 1998, p. 14).

The only available mark of the mental in the extended mind approach concerns the functional analysis of the resource in a given context. The idea is that in the extended mind approach the mark of the mental is not something already given; rather, it is something one discovers, starting from an intuitive and shared idea of what a mental process is. The minimal and operational mark derived from this "commonsense functionalism," however, seems to be just a pragmatic instrument that does not characterize the mental in a substantive manner and is limited to granting a "cognitive" and/or "mental" status to those parts of a system that play a central role in a "recognizably cognitive process."

> These are cases when we confront *a recognizably cognitive process*, running in some agent, that creates outputs (speech, gesture, expressive movements, written words) that, recycled as inputs, drive the cognitive process along. In such cases, any intuitive ban on counting inputs as parts of mechanisms seems wrong (Clark, 2008, p. 131, our italics).

> It is, above all else, a matter of empirical discovery, not armchair speculation, whether there can be a fully fledged science of the extended mind (Clark, 2008, p. 95).

The problem is that such a minimal and operational mark of the mental is very unlikely to save the model from the risk of overextension. We believe that it is necessary to deepen the analysis of what is mental, referring to the role of consciousness and of personal level in depicting an adequate mark. The same attitude is shared by authors that have dealt with the mark of the mental or the problem of criteria (such as Rupert, 2004; Gertler, 2007; Rowlands, 2009; Roberts, 2012; Adams and Garrison, 2013; Varga, 2018), although we do not have the opportunity to discuss them here.

In the next paragraph, we will sketch our own solution, starting from the rediscovery of the role of consciousness in marking the mental and based on the notion of transparency (see Clark, 2004, 2008; Wheeler, 2019). It is developed from some of our previous works on the topic (Di Francesco and Piredda, 2012; Di Francesco et al., 2016; Di Francesco and Tomasetta, 2017) and from a valuable discussion of the mark of the mental by Horgan and Kriegel (2008), recently revisited by Gallagher (2017). Lastly, we will consider some general conclusions concerning the extended mind framework that derive from it.

The Mark of the Mental: Some Preliminary Thoughts

To sum up, we find ourselves in a situation in which the search for Criteria to Avoid Overextension (CAOs) ends in a dilemma. On the one hand, it seems that keeping the first three criteria and rejecting the fourth—the past-endorsement criterion—will open the extended mind framework to a potentially undesired proliferation of extended beliefs. On the other hand, keeping all four criteria has proven problematic for the extended mind model, as it would imply an overly privileged position for consciousness in deciding what counts as a belief—a position not applicable to internal states.

A straightforward alternative to the problem of finding the right criteria, as already mentioned, is to offer a proper mark of the mental. This is, however, no simple task, and many proposals have already been made on the topic. The particular perspective we wish to take on this subject comes from an

on the mark of the cognitive are easily transferable to the issue of the mark of the mental.

[5]By the way, these two quotes represent a very clear example of the relaxed use of "mind" and "cognitive" in the debate about the extended mind. Another example is the following quote from Gallagher: "The strict distinction between causality and constitution is closely tied to the idea that there is a 'mark of the mental' (a way to determine what processes count as cognitive and what processes do not)" (2017, p. 7).

acknowledgment of the importance of the role that the past-endorsement criterion has had to play in this story. We think that the fluctuating presence of the past-endorsement criterion in the literature on the extended mind indicates something interesting about the role it was meant to play. Our contribution to the debate would be to sketch a possible version of the mark of the mental that also has the merit of defining the suspended status of the conscious past-endorsement criterion, thereby establishing an often neglected issue. Before introducing our proposal, some preliminary—though simplified—considerations are in order.

The battlefield of the mark of the mental has been traditionally divided into two areas: on the one hand, broadly following Franz Brentano, it has been claimed that *intentionality* is what mainly characterizes the mental domain. On the other hand, *consciousness* has been considered the distinctive feature of our mind. Now, on which side of the field should a mind-extender line up?

If one goes for the intentionalist side, one has to remember that the distinction between intrinsic and derived intentionality is not necessarily available to the mind-extender (see Searle, 1992; Clark, 2005; Dennett, 2009). Moreover, if one lacks the means to distinguish between the two, it would be difficult to distinguish between natural and artificial intentional systems, as long as they entertain intentional states.

On the other hand, the conscientialist option should be further specified. One could think that phenomenal consciousness is what distinctively characterizes our mental experience, but this is not the only possible interpretation of the role of consciousness in defining a mark of the mental. Consciousness is also a particular way through which we have access to our mental states, one that, at least since Descartes, has played a fundamental role in the construction of theories of mind. We seem to have direct access to our mental states, and we act according to them without questioning whether they are really ours. This condition is something very similar to what the "glue and trust" criteria—along with the conscious past-endorsement criterion—attempt to grasp. Even if it is implausible to claim that every single mental state—say, a belief—has been consciously endorsed before entering our mind, we believe the reference to the role of consciousness, and the particular way we have access to some contents of our mind, to be nevertheless meaningful. Even if the conscious past-endorsement criterion has to be rejected, its pointing to consciousness may represent an appropriate suggestion to follow.

This is the intuition we intend to follow in the remainder of this paper: to rediscover the central role of consciousness in accounting for the specific features of our mind. In so doing, we will conclude that the past-endorsement criterion is wrong, but that it nevertheless indicates the right direction to follow in acknowledging a fundamental role to consciousness in defining what can count as mental.

Our path will be divided into two steps: the first concerns the form, or the structure, of the mark of the mental; the second regards its content.

Usually, when we think of the form of the mark of the mental, we imagine a feature or a set of features that, if possessed by a process or a state, unmistakably qualify that process or state as mental. They can be considered as necessary and sufficient conditions for mentality. This way of looking at the problem makes the quest for the mark of the mental even more difficult that it already is, as it demands a great deal of any theory of the mental[6]. However, the individuation of necessary and sufficient conditions is not the only possible kind of a mark of the mental and, of the other possible candidates, we will rely on the "two-layer" mark of the mental by Horgan and Kriegel (2008)[7], based on the prototype theory (Rosch, 1973)[8].

According to Horgan and Kriegel (2008), the concept "mental" is organized as a prototypical concept (Rosch, 1973). If this is so, there are some prototypical mental states that constitute the standard cases, and other states that can be defined as mental in virtue of a relationship they entertain with the prototypical cases. In Horgan and Kriegel's view, the prototypical mental states are phenomenally intentional states[9], defined as "uncontroversially, unquestionably, paradigmatically, prototypically mental" and "other mental states count as mental only when, and insofar as, they bear the right relationship to phenomenally intentional states" (p. 8). An interesting aspect of this view is that, depending on the intensity of the relation with the prototypical mental states, the mentality of the other states comes in degrees, admitting "gray areas in which there is no deep fact of the matter as to whether a given state is mental or not." This is the reason why Horgan and Kriegel speak of a "two-layer" mark: the first layer is composed of phenomenally intentional states, "the only ones that qualify as mental in and of themselves and regardless of any relationship they might bear to any other state," while the second layer is composed by all "the relevant states [. . .] that are causally integrated in the right way within larger systems that feature phenomenally intentional states" (p. 10).

Now, while Horgan and Kriegel choose phenomenal intentional states as the prototypical mental states, it is of course possible to select other states as prototypical and still keep the prototypical structure of the mark of the mental. This is what we propose later in this work. The time is now ripe to present our proposal, dedicating some thoughts to the content of the mark of the mental.

[6] As far as we know, the proposals by Adams and Aizawa (2008) and by Rowlands (2009) regarding the mark of the cognitive adopt this intuition, and both have encountered considerable problems.

[7] More recently, Kriegel (2017) has proposed to interpret "mental" as a natural kind concept, having a necessary and sufficient underlying nature. In any case, already in Horgan and Kriegel (2008), fn. 24, p. 370 it is specified that "there is no real tension between being a natural kind concept and being a prototype concept. A natural kind prototype concept would be one for which the relevant relationship non-prototypical instances would have to bear to prototypical ones is that of (probably exact) similarity with respect to underlying nature."

[8] Another possibility is to adapt Gallagher's "pattern theory of the self" to the case of the mark of the mental (Gallagher, 2013, 2017). We find this proposal unattractive as, in our view, more than a theory of the mark of the mental, this would qualify as a theory concerning the non-existence of a mark of the mental, and we would like to believe that—also being a natural/biological category—having a mind could be somehow described in a substantive manner.

[9] "The phrase "phenomenal intentionality" denotes a kind of intentionality that phenomenally conscious states exhibit and moreover exhibit precisely *in virtue* of being phenomenally conscious states, that is, in virtue of their specific phenomenal character" (Horgan and Kriegel, 2008, pp. 5–6).

Sketches for a Transparency-Based Mark of the Mental

In the last section, we specified that in our view the mark of the mental should not be considered as a necessary condition that an agent's mental states have or do not have, but rather as a prototypical concept to which it is possible to be nearer or further (this idea is inspired by Horgan and Kriegel, 2008). Having established this, we can now tackle the question of the content of the mark of the mental.

Our proposal is that (1) conscious states are the prototypical mental states and (2) some unconscious/subpersonal states can also legitimately be considered as mental: this happens when they have a particular relation with conscious states (Di Francesco and Piredda, 2012; Di Francesco et al., 2016; Di Francesco and Tomasetta, 2017). Condition (2) will also apply for some extended putative mental states.

The question is now: how is such a relation to be specified? We suggest that the main characteristic to describe this relation is "transparent access." We rely on the conception of transparency developed by Clark (2004, 2008) and Wheeler (2019), inspired by the phenomenological tradition (see Heidegger, 1927; Merleau-Ponty, 1945). This is a conception of "phenomenological transparency" in the sense that it depends on what is perceived and experienced by the agent. In the famous example by Heidegger, the skilled carpenter has no conscious recognition of the hammer in use: "when we skilfully manipulate equipment in a hitch-free manner, we have no conscious apprehension of the items of equipment in use as independent objects, that is, as something like identifiable bearers of determinate states and properties" (Wheeler, 2019, p. 859). Tools in use become thus phenomenologically transparent. Speaking of the body, Clark writes:

> At such moments, the body has become "transparent equipment" (Heidegger, 1927/1961): equipment (the classic example is the hammer in the hands of the skilled carpenter) that is not the focus of attention in use. Instead, the user "sees through" the equipment to the task in hand. When you sign your name, the pen is not normally your focus (unless it is out of ink etc.). The pen in use is no more the focus of your attention than is the hand that grips it. Both are *transparent equipment*. (Clark, 2008, p. 10, our italics)

This conception of transparency is thus construed in analogy with the transparency in tool use and in technology. In this context, a process (even an "extended" process) is taken to be transparent if it is invisible to the subject, who uses it in a fully unconscious and automatic way; yet, the results of the process must be accessible to the subject's consciousness (even if the process itself is not). In this way, we achieve a strengthening of the link between the mental and the conscious (Di Francesco and Piredda, 2012, Chap. 5).

Our idea is to take transparent access to consciousness as fundamental for mentality: being transparently accessible by consciousness, or being sufficiently integrated with a mental state which is transparently accessible by consciousness, rather than being internal to the skull, is what makes something mental. In this sense, transparency expresses the idea of a strong integration between the subject's conscious mind and her other mental processes—where integration is to be considered a relation of coupling in which a component's output is recycled as input from the other component—as in the case of the output of an unconscious process that is used as input from a conscious one.

> There is no special magic associated with direct physically wired links between components. The differences between links forged by nerves and tendons, by fiber-optic cables, and by radio waves are relevant only insofar as they affect the timing, flow, and density of informational exchange. These latter factors are relevant, in turn, because they affect the nature of our relationship with the various kinds of tools, equipment, and subsystems. If the links are sufficiently rich, fluid, bidirectional, fast, and reliable, then the interface between the conscious user and the tool is liable to become transparent, allowing the tool to function more like a proper part of the user. (Clark, 2003, p. 103)

Transparency brings about a sort of direct access of the subpersonal content to consciousness—in the sense that at the phenomenological level the given content is directly available to the conscious/personal mind of the subject. In other cases, transparency plays a less direct but still relevant role:

> Applied to the mark of the mental issue, this allows us to regain, for instance, those subpersonal states that are seemingly endowed with a representational content (e.g., Marr's $2^1/_2$-D sketch, or perceptual processing in the ventral pathway) and, though being not directly accessible by the personal mind, are sufficiently integrated with personal processes. In sum, (derived) mentality requires integration between conscious and unconscious. (Di Francesco et al., 2016, p. 46)

According to this view, Marr's $2^1/_2$-D sketches represent a good example of how integration—together with the transparency of the final output—may drive the individuation of derived *internal* mentality. Analogously, there may be extended processes that involve states or processes that, though not strictly transparent themselves, can be considered as cases of derived *extended* mentality in virtue of their being strongly integrated with other (extended, mental) processes. Examples of this kind could be the processes Otto uses in order to retrieve his extended beliefs on the notebook and some processing of an external cognitive prosthesis at work (see Vold, 2015, pp. 26–27).[10]

The fact that we believe that the prototypical mental cases are conscious states is not to say that *only* conscious states are mental—which would be a *strong* conscientialist proposal—but just to submit that mental states, conscious or unconscious, should stay in the right sort of relation to the personal/conscious mind. For this reason, we have qualified our proposal as "weak conscientialism."

Interestingly, a mark of the mental based on the degree of transparency offers the possibility of providing a continuous and somehow measurable mark. So it could be possible, in principle, to elaborate a "scale of mentality" based on a

[10] Another interesting case is the one of "language scaffolding," when for example we are writing a paper and we rely on a series of external resources in order to get our job done. Some of these resources could be considered transparent and some others not, but they are nevertheless so intimately integrated into the extended cognitive process to be considered also part of the extended mental system (see Clark, 1997, pp. 206–207). We would like to thank a reviewer for having pushed us to make clear the role of integration, together with transparency, not only in the cases of non-conscious *internal* mentality but also in the cases of *extended* mentality.

multidimensional matrix. One proposal in this sense has been advanced by Heersmink (2012) concerning mind–artifact relations. The dimensions considered include reliability, durability, trust, procedural and representational transparency, individualization, bandwidth, speed of information flow, distribution of computation, and cognitive and artifactual transformation. In this way, the concept of mentality could be considered a nuanced and more inclusive concept. Of course, depending on how far one is willing to stretch the concept in the direction of less prototypical cases, the concept can—to greater or lesser degrees—be kept in line with our intuitive comprehension of it.

A last important point regards how our criterion works in several examples of putative extended mentality. In particular, we would like to test it in two cases: the already mentioned case of the extremely efficient electronic phone book imagined by Rupert (2009) and previously mentioned this paper and the case of some contents accessible through Google, often used as a possible counterintuitive consequence of the extended mind framework.

According to our reasoning, we believe that, in the case of the very efficient electronic phone book, we could consider its contents as plausible examples of extended beliefs if the first three criteria indicated by Clark and Chalmers are satisfied and if the electronic phone book is perceived by the agent as a transparent resource. This feeling of transparency can be continuous, or it may change over the course of the agent's life—it is possible that when we buy a new electronic tool, for example, there is a transition period during which we are more familiar with the old tool, but then we gain familiarity with the new one, which "magically" becomes transparent. Thus, if the instrument "disappears" when we use it, it is legitimate to consider it as a piece of our extended mind.

The case of the contents of Internet pages accessed through Google is entirely different in our view. In fact, even if we could imagine the day in which we can access Internet pages by wearing a pair of Google glasses or in other very immediate and direct ways, there still is a criterion that seems not to be satisfied by this kind of resource: that is, the automatic endorsement, according to which the agent (say, Otto), upon retrieving information from the notebook, automatically endorses it. It is very unlikely to imagine that we would endorse any possible content transparently and immediately retrievable from the Internet, and this is a good reason—at least for the moment—to leave Internet pages out of our extended mind.

So, in conclusion, what our criterion of "transparent access" adds to the first three criteria is a phenomenological condition on how we "live" our relation with the resource in question. It is a functional–phenomenological condition, quite far from the historic condition proposed by Clark and Chalmers.

As we have seen, the notion of transparency we have in mind has several components: immediacy, direct availability, and integration, and in our view it should help us discriminate mental from non-mental resources. The reference to transparent access to mark the mental is useful to discriminate the mental from the non-mental from both sides: from the inside, to distinguish subpersonal states that "deserve" the label "mental" (for example Marr's 2 $^1/_2$–D sketches) from states that are very far from the mind (e.g., low-level neurophysiological states); from the outside, to distinguish plausible cases of extended mental states (for example, Otto's extended mental states) and less plausible ones (e.g., transparently retrievable contents of any Google page).

By putting these two steps (form and content of the mark of the mental) together, we are able to sketch a solution to the problem of criteria and the mark of the mental. On the one hand, the notion of mental can be extended to incorporate subpersonal phenomena, provided that these are somewhat integrated with conscious processes (as shown by the concept of transparency). On the other hand, in accordance with this criterion for the mental, we claim that the subpersonal approach has to be integrated with reference to the personal level, in contrast with the approaches that fail to appreciate the link between personal and subpersonal.

CONCLUDING REMARKS

The weak conscientialist mark of the mental we have just sketched, which gives a central role to consciousness and personal mind, seems to concede much to an internalistic picture of the mind. From this point of view, our proposal seems to share with Horgan and Kriegel (2008) and Farkas (2012) the downplaying of the philosophical significance of the extended mind hypothesis. This might be true at the metaphysical level (the paradigm shift imposed by the extended mind hypothesis does not affect the centrality of the personal mind), but on the methodological and anthropological levels, things are different. On the methodological side, only time will tell what progress an externalist investigation of mental states can provide. On the anthropological side, the consideration of human beings as natural-born cyborgs can lead us to review our vision of human beings, with evident philosophical and ethical follow-ups.

In particular, we think that the significance of the extended mind model is not limited to the metaphysical or epistemological evaluation of the mental (that, even by Clark's admission, could be in principle non-provable on empiric grounds, see Clark, 2011). The extended mind model is worth analyzing also for its anthropological and cultural significance: it helps us recognize our fundamental debts toward the external environment in constructing our habitual everyday lives. We think that acknowledging our nature as "natural-born cyborgs" (Clark, 2003) helps us show that extended cases of mentality should not be considered as such extravagant and uninteresting cases of mentality as Horgan and Kriegel seem to think (2008, p. 22): rather, the important way in which we delegate to external resources so much of our thought and private information testifies to the importance of these material external resources in the construction and maintenance of our thoughts and memories. Moreover, disregarding this phenomenon could represent a serious shortcoming for a contemporary theory of the (extended) mind.

Our solution could perhaps be labeled as weakly Cartesian, because of the central role of the conscious mind. However, at the same time it allows moderate extensions of the mind—and paves the way for the philosophical anthropology of the "natural born cyborgs" proposed by Clark (2003)—a view of human nature that

we take as one of the most significant by-products of the adoption of the extended mind stance.

AUTHOR CONTRIBUTIONS

GP and MD conceived this manuscript. GP wrote the first draft of the manuscript. Both authors contributed to manuscript final version, its revision, and they read and approved the submitted version.

FUNDING

This work has been funded by the PRIN Project "The Mark of Mental" (MOM), 2017P9E9N, active from 29.12.2019 to 28.12.2022, financed by the Italian Ministry of University and Research.

ACKNOWLEDGMENTS

We would like to thank MD's coauthors and our colleagues Massimo Marraffa and Alfredo Paternoster. Special thanks go to Alfredo Tomasetta, who gave valuable comments on a previous version of the manuscript. We would also like to thank the audience of the conferences where we presented this work for their helpful comments: Silfs Conference 2017 in Bologna, ECAP Conference 2017 in Munich, and AISC Conference 2019 in Rome. Finally, we thank the two reviewers for their constructive comments and suggestions.

REFERENCES

Adams, F., and Aizawa, K. (2001). The bounds of cognition. *Philos. Psychol.* 14, 43–64. doi: 10.1080/09515080120033571

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

Adams, F., and Garrison, R. (2013). The mark of the cognitive. *Minds Mach.* 23, 339–352.

Chalmers, D. (2019). "Extended cognition and extended consciousness," in *Andy Clark and His Critics*, eds M. Colombo, E. Irvine, and M. Stapleton (Oxford: Oxford University Press).

Clark, A. (1997). *Being There. Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. (2003). *Natural-Born Cyborgs. Minds, Technologies, and the Future of Human Intelligence*. New York, NY: Oxford University Press.

Clark, A. (2004). Author response, in we have always been . . . cyborgs. *Metascience* 13, 169–181.

Clark, A. (2005). Intrinsic content, active memory and the extended mind. *Analysis* 65, 1–11. doi: 10.1093/analys/65.1.1

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

Clark, A. (2009). Spreading the Joy? Why the Machinery of Consciousness is (probably) still in the Head. *Mind* 118, 963–993. doi: 10.1093/mind/fzp110

Clark, A. (2010a). *Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa*, ed. R. Menary (Cambridge, MA: MIT Press), 81–99.

Clark, A. (2010b). *Memento's Revenge: The Extended Mind, Extended*, ed. R. Menary (Cambridge, MA: MIT Press), 43–66.

Clark, A. (2010c). Much Ado About Cognition. Reviewed Work(s): The Bounds of Cognition by Frederick Adams and Kenneth Aizawa; Cognitive Systems and the Extended Mind by Robert D. Rupert. *Mind* 119, 1047–1066. doi: 10.1093/mind/fzr002

Clark, A. (2011). Finding the mind: Book symposium on supersizing the mind: Embodiment, action, and cognitive extension. *Philos. Stud.* 152, 447–461. doi: 10.1007/s11098-010-9597-x

Clark, A. (2012). Dreaming the whole cat. Generative models, predictive processing, and the enactivist conception of perceptual experience. *Mind* 121, 753–771. doi: 10.1093/mind/fzs106

Clark, A. (2016). *Surfing Uncertainty. Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

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

Dennett, D. C. (2009). "Intentional systems theory," in *The Oxford Handbook of Philosophy of Mind*, eds B. McLaughlin, A. Beckermann, and S. Walter (Oxford: Oxford University Press), 339–350.

Di Francesco, M. (2007). "Extended cognition and the unity of mind. Why we are not «spread into the world»," in *Cartographies of the Mind*, eds M. De Caro and M. Marraffa (Berlin: Springer), 213–227.

Di Francesco, M., Marraffa, M., and Paternoster, A. (2016). *The Self and Its Defences. From Psychodynamics to Cognitive Science*. London: Palgrave Macmillan.

Di Francesco, M., and Piredda, G. (2012). *La Mente Estesa. Dove finisce la mente e comincia il mondo*. Milano: Mondadori Università.

Di Francesco, M., and Tomasetta, A. (2017). A not-so-extended mind. *Reti Saperi Linguaggi*, 12, 261–273.

Farkas, K. (2012). Two versions of the extended mind thesis. *Philosophia* 40, 435–447. doi: 10.1007/s11406-011-9355-0

Gallagher, S. (2013). A pattern theory of self. *Front. Hum. Neurosci.* 7:443. doi: 10.3389/fnhum.2013.00443

Gallagher, S. (2017). *Enactivist Interventions. Rethinking the Mind*. Oxford: Oxford University Press.

Gallagher, S. (2018). The extended mind: state of the question. *Southern J. Philos.* 56, 421–447. doi: 10.1111/sjp.12308

Gertler, B. (2007). "Overextending the mind," in *Arguing About the Mind*, eds B. Gertler and L. Shapiro (New York, NY: Routledge), 192–206.

Heersmink, R. (2012). "Mind and artifact: a multidimensional matrix for exploring cognition-artifact relations," in *Proceedings of the 5th AISB Symposium on Computing and Philosophy*, eds J. M. Bishop and Y. J. Erden (Birmingham: AISB), 54–61.

Heidegger, M. (1927). *Being and Time*, trans. J. Macquarrie and E. Robinson (Oxford: Basil Blackwell).

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.

Horgan, T., and Kriegel, U. (2008). Phenomenal intentionality meets the extended mind. *Monist* 91, 353–380.

Kirchoff, M., and Kiverstein, J. (2018). *Extended Consciousness and Predictive Processing. A third-Wave View*. New York, NY: Routledge.

Kriegel, U. (2017). "Brentano's concept of mind: underlying nature, reference-fixing, and the mark of the mental," in *Innovations in the History of Analytical Philosophy*, eds S. Lapointe and C. Pincock (London: Palgrave), 197–228. doi: 10.1057/978-1-137-40808-2_7

Lycan, W. G. (2002). The case for phenomenal externalism. *Nous* 35, 17–36.

Manzotti, R. (2019). Mind-object identity: a solution to the hard problem. *Front. Psychol.* 10:1–16.

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. London: Palgrave Macmillan.

Menary, R. (ed.). (2010). *The Extended Mind*. Cambridge: MIT Press.

Merleau-Ponty, M. (1945). *Phenomenology of Perception*, trans. C. Smith (New York, NY: Routledge).

Piredda, G. (2017). The mark of the cognitive and the coupling-constitution fallacy: a defense of the extended mind hypothesis. *Front. Psychol.* 8:2061. doi: 10.3389/fpsyg.2017.02061

Roberts, T. (2012). Taking responsibility for cognitive extension. *Philos. Psychol.* 25, 491–501. doi: 10.1080/09515089.2011.622361

Rosch, E. H. (1973). Natural categories. *Cogn. Psychol.* 4, 328–350.

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philos. Psychol.* 22, 1–19. doi: 10.1080/09515080802703620

Rowlands, M. (2010). *The New Science of the Mind. From Extended Mind to Embodied Phenomenology*. Cambridge: MIT Press.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *J. Philos.* 101, 389–428. doi: 10.5840/jphil2004101826

Rupert, R. (2009). *Cognitive Systems and the Extended Mind*. New York, NY: Oxford University Press.

Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.

Sutton, J. (2010). *Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process*. ed. R. Menary (Cambridge, MA: MIT Press), 189–225.

Varga, S. (2018). Demarcating the realm of cognition. *J. Gen. Philos. Sci.* 49, 435–450. doi: 10.1007/s10838-017-9375-y

Vold, K. (2015). The parity argument for extended consciousness. *J. Conscious. Stud.* 22, 16–33.

Wheeler, M. (2010). *In Defense of Extended Functionalism*. ed. R. Menary (Cambridge, MA: MIT Press), 245-270.

Wheeler, M. (2019). The reappearing tool: transparency, smart technology, and the extended mind. *AI Soc.* 34, 857–866. doi: 10.1007/s00146-018-0824-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 15 July 2020
doi: 10.3389/fpsyg.2020.01597

How Abstract (Non-embodied) Linguistic Representations Augment Cognitive Control

Nikola A. Kompa[1]*and *Jutta L. Mueller*[2,3]

[1]Institute of Philosophy, University of Osnabrück, Osnabrück, Germany, [2]Psycho/Neurolinguistics Group, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany, [3]Department of Linguistics, University of Vienna, Vienna, Austria

Recent scholarship emphasizes the scaffolding role of language for cognition. Language, it is claimed, is a cognition-enhancing niche (Clark, 2006), a programming tool for cognition (Lupyan and Bergen, 2016), even neuroenhancement (Dove, 2019) and augments cognitive functions such as memory, categorization, cognitive control, and meta-cognitive abilities ("thinking about thinking"). Yet, the notion that language enhances or augments cognition, and in particular, cognitive control does not easily fit in with embodied approaches to language processing, or so we will argue. Accounts aiming to explain how language enhances various cognitive functions often employ a notion of abstract representation. Yet, embodied approaches to language processing have it that language processing crucially, according to some accounts even exclusively, involves embodied, modality-specific, i.e., non-abstract representations. In coming to understand a particular phrase or sentence, a prior experience has to be simulated or reenacted. The representation thus activated is embodied (modality-specific) as sensorimotor regions of the brain are thereby recruited. In this paper, we will first discuss the notion of representation, clarify what it takes for a representation to be embodied or abstract, and distinguish between conceptual and (other) linguistic representations. We will then put forward a characterization of cognitive control and examine its representational infrastructure. The remainder of the paper will be devoted to arguing that language augments cognitive control. To that end, we will draw on two lines of research, which investigate how language augments cognitive control: (i) research on the availability of linguistic labels and (ii) research on the active usage of a linguistic code, specifically, in inner speech. Eventually, we will argue that the cognition-enhancing capacity of language can be explained once we assume that it provides us with (a) abstract, non-embodied representations and with (b) abstract, sparse linguistic representations that may serve as easy-to-manipulate placeholders for fully embodied or otherwise more detailed representations.

Keywords: embodiment, abstract representations, inner speech, cognitive control, labels

OPEN ACCESS

Edited by:
Leon De Bruin,
Radboud University Nijmegen, Netherlands

Reviewed by:
Guy Dove,
University of Louisville, United States
Anna M. Borghi,
Sapienza University of Rome, Italy

***Correspondence:**
Nikola A. Kompa
nkompa@uni-osnabrueck.de;
nkompa@uos.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *17 March 2020*
Accepted: *15 June 2020*
Published: *15 July 2020*

Citation:
Kompa NA and Mueller JL (2020) How Abstract (Non-embodied) Linguistic Representations Augment Cognitive Control. Front. Psychol. 11:1597. doi: 10.3389/fpsyg.2020.01597

INTRODUCTION

According to embodied approaches to language, comprehending a phrase or utterance requires that one activates embodied, modality-specific – as opposed to abstract – representations. At the same time, evidence is accumulating that language scaffolds impressive cognitive achievements. Language is said to enhance core cognitive functions such as memory, learning, or cognitive control. Yet, the notion that language augments cognition does not square well with embodied approaches to language processing as many explanations of how language enhances cognitive functions draw on the notion of an abstract representation[1]. Unfortunately, neither the notion of an embodied representation nor the notion of an abstract representation is particularly clear; an adequately thorough explication is missing. Moreover, the way in which language augments cognition is not very well understood either; what features of language prove beneficial?; and what are the underlying mechanisms? In what follows, we will first explicate the two notions of representation (abstract vs. embodied) and then examine how language might augment cognition. In order to move the problem onto more tractable ground, we will focus on the way in which the availability and use of a linguistic code enhances cognitive control; thus, all claims we make are confined to that domain. To that end, we will not only discuss embodied and abstract representations but also other types of "linguistic" representations and argue that the cognition-enhancing capacity of language is best explained on the assumption that it provides us with different types of abstract, sparse representations that can generate structure, reduce cognitive load, and increase computational power.

These considerations are primarily meant as a contribution to the debate on language and embodied cognition. But, although not directly addressing the controversial question of whether we think in natural language or whether natural language is constitutively involved in (some forms of) cognition (cf. e.g., Carruthers 2002, for review and discussion), they may nonetheless contribute to that debate as well, as the idea of linguistic representations underlying overt and inner speech may suggest a way in which to spell out the idea that cognition is – in some cases and to some extent at least – linguistic.

REPRESENTATIONS

Embodied and Abstract Representations

The notion of representation or idea (as it was variously called) is notoriously hard to spell out. It has a long and distinguished history. While in antiquity (especially in Plato's work), ideas were not meant to be subjective mental contents but rather immutable, ideal entities, in late antiquity, and the middle ages they began their career as key notions in semiotic and epistemological theorizing. When in modernity (Descartes is often said to be the founding father of modern representationalism), the notion of representation or idea became the heavy-duty notion that we are familiar with today (Perler and Haag, 2010), it was already afflicted with a few problems. Early on, the notion of idea was ambiguous, as ideas were commonly taken to be the vehicles as well as the contents of thought. Also, the notion of abstract idea favored by Locke and others did not fit in with a pictorial conception of idea. Yet, what could a plausible non-pictorial conception of idea (that nonetheless explains how ideas can be acquired in experience) look like? These (and many more) problems have been inherited by recent accounts.

Consequently, some contemporary authors claim that we can (and ought to) do without a notion of representation (understood as something that matches the content of conscious experience) altogether (cf. e.g., Noë and Thompson, 2004). Others still maintain the notion of abstract and thus amodal representation as a core ingredient of mental computations (Dove, 2009, 2011, 2016; Binder, 2016)[2]. This, in turn, is challenged by those who claim that abstract (amodal) representations are dispensable; all it takes are embodied (modality-specific or perceptual) representations (cf. e.g., Prinz 2002; for early critiques cf. Machery 2006, 2007; Mahon and Caramazza 2008). The latter debate is often framed in terms of how concepts or conceptual knowledge may be represented in the brain (Mahon and Hickok, 2016).

With this debate as a starting point, we will first explore what types of representation underlie language production and comprehension. Importantly, we will distinguish between *conceptual* (embodied and abstract) representations (encoding conceptual information) and *linguistic* representations that form the linguistic code itself at its various levels. In line with common usage (within philosophy and psychology), we take conceptual representations to encode (semantic) information about concepts (or categories) such as DOG or CHAIR. But then, some linguistic representations encode *conceptual-linguistic* information, i.e., information about linguistic concepts or categories (VERB, NOUN, and so on). Thus, strictly speaking, one ought to distinguish between what one might call "conceptual-semantic" and "conceptual-linguistic" representations. To keep things as simple as possible, we will nonetheless continue to speak of "conceptual representations" and "linguistic representations" (encoding linguistic information of various types, cf. section Linguistic Representations), unless more detail is required. The aim is to better understand the ways in which language and the various types of representations it affords us may prove cognitively beneficial and may be engaged during cognitive tasks. Eventually, we will argue

[1]The position defended here is incompatible with embodied accounts that claim that linguistic processing necessarily and predominantly recruits sensory-modal areas in the brain. More moderate, hybrid accounts that acknowledge a role for abstract representations and allow for more flexible activation of different types of representations in linguistic processing are compatible with our view (see section Discussion and Open Questions).

[2]Dove, in fact, defends a hybrid view, arguing that we need embodied as well as dis-embodied representations (Dove, 2011). He has it that language gives us "access to a new type of representational format" (Dove, 2014, p. 373). It is "an external symbol system – one that has the computational features associated with amodal symbol systems – that we learn to manipulate in an embodied and grounded way" (Dove, 2018, p. 1).

that language, by providing us with various kinds of representations, enhances different cognitive functions (in particular cognitive control) in different ways. Therefore, in what follows, we will employ a rather thin and uncommitted notion of representation. A representation, as we will use the term, is a pattern of neural activity that fairly robustly encodes information and is thus sensitive to that type of information (which may range from concrete sensory input to generalizations over or abstractions from such input); furthermore, its cognitive role is (at least in part) grounded in the fact that it encodes this information.

It is a well-rehearsed point in the literature by now that sensorimotor areas are activated during language processing (cf. Meteyard et al., 2012; Mahon and Hickok 2016, for recent reviews). It has been observed, for example, that when people listen to sentences or phrases containing action verbs (such as "grasp" or "kick"), motor areas are activated (cf. e.g., Hauk et al., 2004; Pulvermuller, 2005; Aziz-Zadeh et al., 2006). More specifically, roughly the same region is activated when hearing the phrase "grasping the pen" and when seeing a video of a hand grasping a pen (Aziz-Zadeh et al., 2006). Barsalou speaks of "neural reuse" (Barsalou, 2016, p. 1129–1130) in these cases (cf. also Barsalou, 1999)[3]. It is claimed that in coming to understand a particular term or sentence, a prior experience has to be simulated or reenacted. The representation thus activated is embodied (modality-specific) insofar as specific sensorimotor regions of the brain are thereby activated (cf. e.g., Jirak et al., 2010, for review).

Yet, what exactly makes a representation embodied? That it is in a sensorimotor format, some say (Mahon, 2015; Mahon and Hickok, 2016). Yet, this notion raises further questions.

1. It raises the "important question of whether a simulation is sufficiently fine-grained to merit being called "embodied" rather than being some sort of an abstraction, even if that abstraction is originally grounded in a specific action or situation (Sanford, 2008, p. 189)." How much of an experience has to be embodied or simulated; how detailed does the simulation have to be? And, is not any sensorimotor representation an abstraction already (Mahon and Hickok, 2016)?
2. What exactly is the claim at issue? Is the claim that embodied representations are necessary (or even sufficient) for coming to understand a particular term (or grasping a particular concept)? Or is it rather the claim that embodied representations facilitate comprehension without being strictly necessary?[4] Alternatively, some claim that they are simply epiphenomenal, mere by-products of linguistic or conceptual processing. Patient studies showing dissociation between concept possession (or linguistic comprehension) on the one hand and sensorimotor skills on the other speak against too tight a link between embodied representations (i.e., sensorimotor activation) and linguistic/conceptual understanding (cf. e.g., Mahon and Hickok 2016 for discussion). This suggests another, closely related question.
3. Are conceptual representations static and uniform or are they composed differently (or are different types of information drawn upon in a task-sensitive manner) across the variety of situations in which they are activated (Schyns et al., 1998; Vigliocco et al., 2004; Dove, 2016; Mahon and Hickok, 2016; Yee and Thompson-Schill, 2016)? An answer to this question depends on what we mean by "linguistic understanding." The role of embodied representations in linguistic understanding can be adequately adjudicated only against the background of a theoretically sound model of what language understanding amounts to. An example may illustrate the point. Does a congenitally blind person grasp the meaning of the term "yellow" (or possess the concept *yellow*) in a similar manner as a normally sighted person? An answer to that question requires that we specify when understanding is achieved. If we agree that one understands a term if one is able to use it competently in different contexts, to draw valid inferences and to make correct judgments involving it, then we ought to answer the question in the positive (Saysani et al., 2018; Bedny et al., 2019). If, on the other hand, we require that a previous color experience is reenacted or simulated, then we ought to answer in the negative. But then, is not it the point of language that it allows speakers to acquire knowledge that goes beyond immediate sense experience, one might wonder (Dove, 2009, 2014; Binder, 2016)?

While the notion of embodied representation and its role in explaining language comprehension and production invites tricky questions, the notion of an abstract representation is no less problematic. For what is it for a representation to be abstract?

1. On a traditionally influential account, an idea becomes abstract by omission of distinguishing detail (Locke, 1979), thus by compressing information. For example, on seeing various persons, one abstracts from those aspects in which they differ and focuses only on what they have in common, thereby arriving at the (abstract) idea of a human being. As was already noted by contemporaries, this does not square well with a pictorial conception of ideas, as pictorial representations cannot omit detail *ad libitum*. The notion seems to fit better with a conception of ideas as lists of defining features. Yet, the presupposition that such a list is to be had for every idea seems problematic too. On a more recent account, abstraction is conceived of as transformational invariance, i.e., an increasing tolerance to slight transformations in the input (Buckner, 2018). This characterization is promising as it goes some way toward a functional characterization of what an abstract representation

[3]The idea of neural reuse is developed in detail by Anderson, who suggests that "the brain achieves its variety of function by using the same regions in a variety of circumstances, putting them together in different patterns of functional cooperation" (Anderson, 2014, p. 5; cf. also Anderson, 2010).

[4]These two options do not exhaust the space of possibilities. Embodied representations could be causally relevant without being causally necessary, as something else might play the causal role too. Moreover, embodied representations could be constitutive of comprehension (in that they would have to figure in a mechanistic explanation) as opposed to being causally necessary (comprehension might be counterfactually dependent on embodied representations).

is. It tells us that the more abstract a representation is, the more it will tolerate somewhat transformed inputs.

2. But then, different types of abstract terms – and concepts or representations respectively – ought to be distinguished (Kompa, 2019). While every sort of classification requires abstraction, some terms (such as "red") require that objects be classified according to sensory, *determinate* features. Others require that objects (or events) be classified according to *determinable* features; for example, the term "object" is applicable to entities that all have *a* shape, though not necessarily *the same* shape. Still others require that entities be sorted according to functional or defining features ("tool"), or according to evaluative features ("good"). And, still others require that entities be sorted according to structural or relational features ("being the same as") or higher-order relational features, as when one judges that two pairs of objects have the same first-order relational property. While in all these cases, abstraction may be conceived of as increasing tolerance to transformations of the input in each there are certain features that can vary yet others that have to remain fixed. Most importantly, it is not just "simple," determinate features (such as being a particular shade of red) that need to remain fixed but increasingly "complex" features manifesting a certain relational or evaluative structure. Unsurprisingly, then, the process of abstraction is often thought to result in hierarchies of increasingly abstract or complex representations. Also, integration and abstraction seem to go hand in hand. Mastery of evaluative terms, for example, consists in tolerating variation in the input while integrating information about a system of values and norms.

   Early on, proponents of embodied accounts of concept representation discussed abstract concepts (Barsalou, 1999), carefully examined the content of abstract terms (Barsalou and Wiemer-Hastings, 2005), and increasingly stressed the diversity and heterogeneity of abstract terms (Borghi et al., 2018b, 2019), resulting in multiple representations views such as the words-as-tools (WAT) model (Borghi et al., 2019). Different types of abstract concepts (for mental, emotional, or metacognitive states, mathematical or physical entities, etc.) are distinguished and said to rely on different cognitive mechanisms (Borghi et al., 2018b; Desai et al., 2018). Barsalou and others also increasingly stress the need for different types of representations of conceptual information (Pulvermüller, 2013; Barsalou, 2016), including abstract or general representations. At the same time, more hierarchical models of abstract representations which are said to "arise from a process of hierarchical conjunctive coding" (Binder, 2016, p. 1098), i.e., exhibit sensitivity to particular combinations of inputs (ibid), are suggested. Furthermore, more and more authors emphasize the role of language in the processing and acquisition of abstract concepts (cf. e.g., Barsalou et al., 2012, who stress the role of linguistic forms; Borghi et al., 2018a, who briefly address the role of inner speech; or Lupyan and Winter, 2018, who discuss labels and the role of (a lack of) iconicity). For all that, a thorough and systematic account of different types of abstract concepts, the ways in which they are abstract as well as the cognitive infrastructure underlying their mastery is still pending.

3. Most importantly, representations (underlying language processing) can encode not only conceptual-semantic information (in a more embodied or more abstracted fashion) but also other types of linguistic information. They can, for example, encode – and be sensitive to – morpho-syntactic or phonetic information (cf. section Linguistic Representations). That may result in rather sparse, abstract, easy-to-compute representations which can act as placeholders or stand-ins for (maybe even as a sort of pointer to) more detailed, richer representations[5]. They could be thought of as a sort of interface that encodes only very little information itself but can activate associated (sensory-motor, evaluative, affective, etc.) information in a task-sensitive manner. Also, they may invite combination and can help generate structured representations.
4. Finally, one might distinguish abstract (amodal) from multimodal representations. While the former would be responsive to slightly transformed inputs, the latter would be responsive to inputs from various modalities (Fernandino et al., 2016). Thus, multimodal representations might share features with abstract representations, such as integration and tolerance to transformations in the input, and also share features with embodied representations by encoding highly modality-specific, concrete information.

In sum, abstract representations encode less detail than embodied, modality-specific representations. They may be sensitive to relational and otherwise more complex, abstract properties and tolerate various transformations of the input. And, they may be sensitive to different types of information. Being abstract and sparse, they ought to increase computational efficiency and come with low transfer costs (Machery, 2016) as well as help to avoid cognitive load problems (Dove, 2011). Most importantly, conceptual (semantic) representations (be they embodied or abstract) are not the only representations involved in language processing – a fact that is not sufficiently acknowledged in current debates on the topic, or so we will try to show. There are different types of linguistic representations, which encode – and are thus sensitive to – various types of information.

Linguistic Representations

All models of language and language processing assume different types and levels of linguistic representation. Linguistic representations can encode sensory-motor as well as more abstract information. Many linguistic theories differentiate between two mental systems that are involved in language processing, i.e., the mental lexicon, as a storage system for words and the mental grammar, as the set of rules that specify how linguistic units are combined (Bloomfield, 1933; Chomsky, 1965; Garrett, 1976; Pinker, 1991).

[5]Of course, we do not originate this idea (cf. e.g., Barsalou 2016, p. 1134 for a brief review); yet often, existing accounts do not bother to spell out in detail what they mean by "linguistic representation."

Others see in this dichotomy merely a descriptive tool and suggest dropping a strict two-system view when it comes to investigating the functional processes that form the basis of language (e.g., Jackendoff, 2007). As we are interested in the types of representations afforded to us by language, we will not take sides in this debate. Different theoretical accounts make different assumptions about the content and the functional and anatomical substrate of different representations and how different representational levels interact. Yet, there is a basic agreement that sound-level representations (phonology), representations of syntactic classes and operations (syntax), and representations of meaning or concepts (semantics) ought to be distinguished (Chomsky, 1957; Levelt, 1989; Jackendoff, 2007). Thus, it seems clear that language provides us with one or more levels of representation (in addition to conceptual-semantic representations) that could potentially feed into various cognitive processes.

First, as language is coded by sound (or script), there is a level representing the sensory content of the linguistic unit, i.e., phonetic or orthographic information. As speech sounds and their combinations are perceived in a language-specific manner (Miyawaki et al., 1975; Massaro and Cohen, 1983; Werker and Tees, 1984), we must have stored representations of phonemes and phonotactic regularities of our language(s). Further, it is assumed that there is *at least* one intermediate level of lexical representation between the representation of the single sounds and the conceptual level. Most models in fact assume two levels, the lemma level of representation, in which syntactic properties and meaning are specified (Levelt, 1989), and the lexeme level, in which the specific phonological form is laid down (Dell, 1986; Levelt, 1989; Caramazza and Miozzo, 1997). Whether there is an additional lemma level of representation assumed or not, it is uncontroversial that meaning-related, syntactic and phonological information about linguistic units can be accessed independently (Caramazza and Miozzo, 1997; Miozzo and Caramazza, 1997; Roelofs et al., 1998). Those models that do not assume a lemma level directly link semantic and syntactic information to the lexeme level (Caramazza and Miozzo, 1997; Miozzo and Caramazza, 1997). Despite these theoretical disagreements, it seems warranted to assume, following Jackendoff, that words are typically linked to phonological, syntactic, and conceptual-semantic levels of representation (Jackendoff, 2017, p. 193) as is illustrated for the word *cat* here:

- Phonology: /kæt/
- Syntax: +N
- Semantics: CAT

Note that those levels of representation might be very different from each other with respect to the richness, diversity, and structure of their content. Yet, for the current purpose, it is only important that they can be distinguished from each other and not so much how they are precisely characterized.

If we now adopt a view of representations as dynamic entities that are custom-built in a task dependent manner, it seems plausible to assume that language has the potential to provide its users with very different types of representations, depending on the task at hand. At times, this representational code may be sparse and stripped down to, e.g., morpho-syntactic information; at other times, it may be rich and include a whole wealth of conceptual-semantic (and maybe even pictorial or affective) information. More specifically, while lemma or morpho-syntactic representations are, presumably, rather on the abstract side (and while articulatory or motor representations are on the embodied side), phonological representations may be more or less abstract, depending on how much transformation in the input they tolerate. Also, this representational code may benefit from syntactic properties, which makes it easy to combine linguistic units into large and complex (relational) structures, supporting similar structures in other cognitive domains.

In the remainder of this paper, we would like to argue that these properties of representations afforded by language augment other domains of cognition, specifically cognitive control.

COGNITIVE CONTROL

Cognitive control (also termed executive functions) is an important set of processes in the service of optimizing behavior (Cohen, 2017, p. 16). It "is required for adaptive, goal-directed behaviors to solve novel problems, particularly those calling for the inhibition of automatic or established thoughts and responses" (Carlson and Beck, 2009, p. 163). At the very least, it comprises (cf. Cohen 2017 for an overview, and the contributions in Egner 2017 for some details):

a. the ability to detect conflict and to resolve it through various gating mechanisms which result in the inhibition of prepotent, automatic responses.
b. the ability to form, maintain, switch between and update internal goal representations in a task-sensitive manner.

While cognitive control is a well-established construct in psychology, its underlying mechanisms are still subject to debate. Neuropsychological, neurophysiological, and functional imaging research have associated cognitive control with the functions of the prefrontal cortex (Miller and Cohen, 2001). The type and interrelatedness of sub-functions, how exactly cognitive control is represented and computed in the brain, and the representational code of control signals are some of the questions still pending. Theories about cognitive control either focus on unifying, overarching principles, or on the distinctiveness of its sub-functions. The former assume domain-general, uniform principles explaining how various levels of cognitive control are supported by hierarchically organized operations of the prefrontal cortex (Christoff and Gabrieli, 2000; Miller and Cohen, 2001; Koechlin et al., 2003; Badre and D'Esposito, 2007). Yet, what these uniform principles might look like and how cognitive control may be supported by different sub-regions of prefrontal cortex along an anterior-to-posterior gradient is a topic of current debate. For example, it has been suggested that the temporal integration window of cognitive control (ranging from immediate stimulus processing to the integration of information about the past

and the future; Koechlin et al., 2003) or the degree of abstraction in hierarchical action representations (Badre and D'Esposito, 2007) underlies the functional distinctions within the prefrontal cortex. Other theoretical approaches focus on the role and relation of the different distinguishable components of cognitive control. Miyake and colleagues (Miyake et al., 2000; Miyake and Friedman, 2012), for example, argue for the distinctiveness of three basic cognitive control operations, i.e., updating, flexibility, and inhibition. Barkley (2001), on the other hand, singles out non-verbal and verbal working memory, self-regulation of emotion, and reconstitution (i.e., flexibility) as core components of cognitive control. Arguably, both types of theories (those focusing on unifying principles and those focusing on sub-functions and domain-specific processes such as cognitive control in the language domain, cf. section The Availability of Labels as Facilitators of Cognitive Control) will help to improve our understanding of the neurocognitive organization of cognitive control (e.g., Jeon and Friederici, 2015; Badre and Nee, 2018).

The types of abstract representations that are accorded a role in hierarchical models of cognitive control are various and often hard to separate from each other by empirical means. Abstractness of representations with regard to prefrontal cortex function is said to result from: (i) domain generality (as opposed to domain specificity), (ii) relational complexity (indicating whether a response has to be sensitive to simple stimulus properties, to first-order, or higher-order relational properties), (iii) temporal abstraction (with response-selection being based on cues relating to different time scales), or (iv) generalization or governance (with abstract representations generalizing over or governing sets of more specific representations; Badre, 2008; Badre and Nee, 2018).

This leads us to questions about the representational infrastructure of cognitive control. The variety of domains that implement cognitive control and its efficiency with respect to novel tasks seems to demand a systematic, combinatorial code specifying the current control demands (Cohen, 2017). If such a code exists, it would be highly plausible that it shares some properties with language, specifically its capacity for abstraction and compositionality, as it needs to be able to work over arbitrary and novel content in similar ways. And even if such a general code supporting cognitive control does not exist, one has to consider that cognitive control processes have to deal with representations of various degrees of abstraction and complexity, ranging from motor sequences to the planning of future action goals. Thus, a cognitive control system must have at least the computational capacity to deal with a large degree of variability and abstraction. Biologically-based computational models have provided mechanisms that could, in principle, achieve symbolic-like computations in the regions involved in cognitive control (Rougier et al., 2005; Kriete et al., 2013). In the subsequent sections, we will explore whether and how language as an input code to this system could act as a booster. Some aspects of cognitive control may be uniquely human, as is the capacity for language. Potentially, this may be partly due to the way in which both systems are functionally integrated. In order to argue in favor of that point, we will bring together evidence suggesting that the availability of a linguistic code supports cognitive control and that active use of language, specifically inner speech, serves the same purpose. Crucially, we hold that it is not so much the embodied aspects of language but rather its abstract and combinatorial nature that is primarily responsible for the enhancement of cognitive control.

THE AVAILABILITY OF LABELS AS FACILITATORS OF COGNITIVE CONTROL

In various studies, it could be shown that different cognitive tasks and functions benefit from the availability of a linguistic code, especially from the availability of symbolic labels. Evidence stems from research on categorization, analogical reasoning, learning, memory, and cognitive control (Xu, 2002; Carlson et al., 2005; Lupyan, 2012; Althaus and Westermann, 2016; Doebel et al., 2018; Huang and Awh, 2018; LaTourrette and Waxman, 2019). For present purposes, we will zoom in on the role that symbolic labels may play for cognitive control. Labels can take on at least two roles here, depending on whether language (production) is the domain that has to be controlled or whether language is a means of controlling. In the former case, the cognitive domain that recruits control processes is linguistic. In the latter case, language (labels) represents operational aspects of the task, e.g., participants could use linguistic task cues such as "if red cube on right side of the screen press right button" to enhance performance. In this case, the task domain is visuo-spatial but the cognitive operation receives linguistic support.

We will treat the first role of labels only briefly, although cognitive control is involved in language processing at many levels ranging from language production to sentence comprehension and specific phenomena like code switching (Levelt, 1989; Hagoort, 2005; Bourguignon and Gracco, 2019; Sulpizio et al., 2020). It touches on the question of how closely linguistic and control systems are connected. We will focus here on whether cognitive control, when it is in the service of language-related tasks, is somehow different from cognitive control during non-linguistic tasks. Jeon and Friederici (2013, 2015) systematically investigated this question and compared linguistic and non-linguistic material with comparable affordances of hierarchical control. Participants were presented with hierarchically structured Korean symbols either with or without linguistic explanations. Both task conditions involved the anterior-to-posterior gradient of cognitive control in the prefrontal cortex (Jeon and Friederici, 2013). Yet, hierarchically structured sentences from the native language, i.e., highly familiar linguistic material, were processed by posterior prefrontal cortex (BA 44) only, even at a high level of hierarchical complexity. The authors argue that the high degree of automaticity that is typical of native language processing impacts on how the prefrontal cortex supports processes of hierarchical control (Jeon and Friederici, 2013). Thus, the idea is that the same type of formal (i.e., hierarchical) control demand engages

different brain areas, depending on whether the task is novel or highly familiar (as is the case in natural language processing). This idea is in line with the view that brain areas supporting language processing are separable from those supporting domain general cognitive control (Fedorenko, 2014). Others argue for a more integrative view, in which language and domain general cognitive control are more intimately intertwined (Rouault and Koechlin, 2018; Bourguignon and Gracco, 2019). Differences between automatic and non-automatic language processing are here explained as differences along the temporal axis of cognitive control, whereby highly automatic language processing involves chunking processes within a single (not hierarchically structured) task-set while non-automatic linguistic processes are supposed to involve the generation of successive independent task-sets (Rouault and Koechlin, 2018). In a similar vein, an integrative view of language and cognitive control is supported by the observation that brain regions that are specialized in language processing and those that belong to domain-general control networks are closely linked during cognitive control in language production tasks (Bourguignon and Gracco, 2019)[6].

This brings us to the second role of labels and the question of how access to a linguistic code scaffolds cognitive control in non-linguistic tasks. Many studies using classical cognitive control tasks have revealed that performance is sensitive to the inclusion of symbolic representations in some aspects of the task. Several studies tested children's performance in a reverse contingency task (in which participants have to point to the smaller of two rewards in order to receive the larger one) in the presence or absence of various types of symbolic labels substituting the real rewards. It was found that children performed better when labels were used (Carlson et al., 2005; Apperly and Carroll, 2009). Interestingly, the beneficial effect of labels does not entirely depend on the availability of a linguistic system, as similar effects were found in great apes (Boysen et al., 1996). Currently it is still unclear which property of symbolic labels causes this effect. It has been suggested that labels increase psychological distance in the face of an immediate reward (Carlson et al., 2005) or that they help to formulate alternative response strategies (Apperly and Carroll, 2009).

Another task that requires the inhibition of a prepotent response is the delay-of-gratification task. Participants have to reject an immediate reward in order to receive a larger reward later. This can be tested by either a choice task or a maintenance task. In the former, the delay cannot be influenced any more after the choice while the latter requires the suppression of the immediate reward for a longer period of time. It is long known that directing attention away from the arousing properties of the reward, e.g., by imagining the reward as a picture, makes it easier for young children to resist the immediate reward (cf. Mischel et al., 1989, for review). Some studies using delay-of-gratification choice tasks reported a reversed effect of symbolic labels, namely an increase of choices in favor of immediate rewards, observable in primates and human children (Addessi et al., 2014; Labuschagne et al., 2017). Yet, these results can also be explained as an effect of symbolic distancing: it is hypothesized that experiments with real food or food pictures may overestimate the abilities to tolerate delays in the participant as they might trigger impulsive choices due to the appetitive nature of the stimuli (Addessi et al., 2014). One of the most plausible explanations for the performance in these tasks seems to be that symbolic labels best sever the link between experience (stimulus) and response by activating abstract ("cool") representations that are not too closely linked to the arousing ("hot") aspects of the experience. Note that labels also impact cognitive control beyond delay-of-gratification or reverse-contingency tasks. It has been shown, for example, that 3-year olds benefitted from labeling in other cognitive control tasks, e.g., the dimensional card sorting task or complex visual search tasks (Kirkham et al., 2003; Miller and Marcovitch, 2011) although there are studies which could not replicate such effects (Müller et al., 2008). If labels were to only activate embodied, modality-specific representations, i.e., simulations of prior experiences, no psychological distance would come about. In line with this view, it has been argued that labels can, occasionally, "carry the burden of conceptual processing under a range of circumstances by effectively acting in place of deeper, more detailed representations of referent meaning" (Connell, 2019, p. 1308), especially in tasks that require only "shallow or superficial conceptual processing" (Connell, 2019, p. 1313). It therefore seems plausible that the representations engaged during those tasks are rather abstract. Kharitonova et al. (2009) and Kharitonova and Munakata (2011) provide direct evidence that participants who successfully perform in a switching task apply more abstract representations compared to less successful participants, as the former are also better in generalizing an acquired rule to novel items.

All these findings and considerations point toward a role for abstract linguistic representations as facilitators of cognitive control. If labels have such a role to play, one may expect that linguistic impairments may affect the performance in cognitive control tasks. This seems to be borne out by the available evidence. Aphasic patients and individuals with developmental language impairments have been shown to be somewhat impaired in cognitive flexibility and inhibition tasks (Baldo et al., 2005; Pauls and Archibald, 2016). Note, though, that there are also dissociations between linguistic and non-linguistic tasks in both aphasia and developmental language disorders, clearly supporting the view that language and complex cognition are not to be equated (Fedorenko and Varley 2016; Archibald 2017, for review). Furthermore, evidence from language impairments rather attests to the active (online) use of language in cognitive tasks (as compared to compensatory strategies) than to the internal (offline) availability of a linguistic code, which is difficult to assess in those cases. These strategic uses of language for the purpose of formulation of cognitive control task affordances will be treated in more detail in the following section.

[6]Also, whether cognitive control processes during language processing are language-specific or not, linguistically coded semantic knowledge may provide an additional control system that can be exploited by non-linguistic domains of cognition, termed "semantic control" (cf. Lambon Ralph et al., 2017; Bourguignon and Gracco 2019, for review).

THE USE OF INNER SPEECH IN SUPPORT OF COGNITIVE CONTROL

At the beginning of the 20th century, Vygotski (1986) propagated the notion that inner speech is internalized public speech and retains some features of the latter (e.g., a social aspect), while losing others (e.g., by being compressed). He claimed that when acquiring a language, the child first talks out loud in what he called, following Piaget, "egocentric" speech, and what is today called "private speech." In the course of development, speech is more and more directed at the child themself, and private speech slowly transforms into inner speech. Inner speech is inaudible to others (Alderson-Day and Fernyhough, 2015, p. 931). The speaker "apprehends him or herself to be speaking meaningfully without producing any accompanying sound or appreciable bodily […] movement (Hurlburt et al., 2013, p. 1482)[7]."

Different things go by the name "inner speech" though. Inner speech ought to be distinguished from the auditory imagery of speech (Machery, 2005; Hurlburt et al., 2013; Gauker, 2018) or inner hearing (Fernyhough, 2016), although in conscious inner speaking, one seems to always accompany the other. It ought to also be distinguished from "unsymbolized" thinking (Hurlburt and Akhter, 2008), although there may be a gradient from fully explicit, articulate (if unarticulated) inner speech to compressed, truncated (still language-based) thinking (that is no longer experienced as "speech," lacking a recognizable phonetic profile). One might hypothesize that the latter still activates (something like) lemma representations but no longer activates phonological or articulatory representations.

Moreover, inner speech is put to different uses and serves different ends. It may occur while one is engaged in a cognitively demanding task, as when one is reflecting on a problem, planning an action, or deliberating more generally; it may take the form of an inner monologue or dialogue (Fernyhough, 2009). It may take the form of self-regulatory and also motivational self-talk, as when one preps oneself for a sporting performance (in which one often addresses oneself as "you"; cf. Fernyhough 2016). One engages in inner speech while silently rehearsing something and also when one is daydreaming, letting one's mind wander (Wiley, 2016). It allows us to think about thoughts, being, arguably, "the single most important tool for intentional ascent" (Bermudez, 2018, p. 204); and so on and so forth.

Most importantly for our purposes, there is an ever-growing body of evidence supporting the notion that private or inner speech enhances children's (and to a lesser extent adults') performance in different memory, planning, and problem-solving tasks (Diaz and Berg, 1992; Winsler et al., 2009). Evidence is accumulating that inner speech enhances cognitive flexibility by aiding retrieval and activation of task goals (Miyake et al., 2004).

It has been shown that verbal self-instructions improves performance in switching tasks, especially in children and the elderly (Kray et al., 2008). One of the most famous examples of a switching task is the Wisconsin Card Sorting Paradigm, in which children are asked to sort bivalent cards (e.g., green boats, red boats, green cows, and red cows) first according to one dimension (e.g., shape) and are then asked to sort along the other dimension (color). In a similar vein, switching costs have shown to increase in adults when inner speech is disrupted, e.g., *via* articulatory suppression (Emerson and Miyake, 2003; Miyake et al., 2004). Recently, a role for inner speech in task switching has also been shown in an interference-free setting, *via* electromyographic recordings from the tongue (Laurent et al., 2016). Yet, it is not only flexibility that is affected by overt or coverts verbalization but also other aspects of control tasks such as inhibition (Kray et al., 2009), task maintenance (Saeki et al., 2013), and control focus (proactive vs. reactive control; Kray et al., 2015) have been shown to be modified by task-related verbalizations. While these examples highlight the function of inner speech as an additional tool for coding task-related representations that are used during task processing, there is further evidence that even evaluative and motivational inner speech that does not directly represent the task can enhance performance in classic cognitive control tasks. Gade and Paelecke (2019) found that participants who reported the habitual use of motivational and evaluative inner speech showed less conflict in two classic cognitive control tasks (the Simon and Flanker tasks). Consistent with these findings, recent reviews conclude that inner speech, while maybe not strictly necessary, nonetheless augments different aspects of cognitive control (Cragg and Nation, 2010; Kray and Ferdinand, 2013).

Especially in cognitively demanding tasks requiring high levels of control, inner speech may help to represent task-related information and to retrieve, maintain, update, and manipulate task representations. The linguistic representations provided by language may serve as a good proxy in order to quickly build or modify abstract control representations. If language was such a support system (instead of an integral component of the control system), one should expect a positive impact of language specifically for unpracticed, novel tasks. Language could serve as an important function in formulating task representations but become superfluous once those representations were installed. A recent study by van't Wout and Jarrold (2020) confirms this intuition by reporting articulatory suppression effects during the initial phase of novel task learning and not in a later phase. Support may also come from studies on rapid instructed task learning (RITL, for short), i.e., "the ability to learn task procedures from instruction" (Cole et al., 2013, p. 1), an "especially important form of cognitive flexibility" (Cole et al., 2013, p. 1) and something humans – as opposed to other animals – excel at. And, while language does not seem to be strictly necessary, and although limited RITL-abilities have also been found in monkeys and non-human primates, RITL that employs linguistic means seems to be "the most powerful form" (Cole et al., 2013, p. 3). This may be due to

[7]Empirically investigating inner speech raises tricky methodological questions and seems to call for a methodologically pluralist approach. Unsurprisingly, then, there are neuroimaging and neuropsychological studies examining the neural correlates of inner speech; others devise questionnaires or engage in descriptive experience sampling (cf. Alderson-Day and Fernyhough 2015, for review).

the fact that it increases high-fidelity transmission of task-relevant information. But, again, one might also hypothesize that language not only helps to formulate task instructions in overt speech but also to come up with and maintain (increasingly abstract and less context-bound) task rules in inner speech. Also, integrative models of RITL highlight the combinatorial properties of the representations underlying task learning and the resulting cognitive flexibility (cf. Cole et al., 2013, for discussion), something linguistic (e.g., lemma) representations could deliver.

All in all, it seems that inner speech influences performance in cognitive control tasks through several mechanisms. At times, it may be useful for the representations of task-related aspects. The abstract and sparse linguistic code may aid memory retrieval, maintenance, and manipulation of task representations. Such computational benefits are easier to explain when taking into account the combinatorial properties of (abstract) linguistic representations (as opposed to embodied ones). At other times, when inner speech improves performance as motivational self-talk, the psychological distancing function of language may come to the fore with inner speech also helping to monitor one's performance and to ensure that one stays on task. Thus, there is probably no unitary function of inner speech that improves cognitive control but rather several aspects of it that, nonetheless, all serve to enhance the uniquely human power of cognitive control.

DISCUSSION AND OPEN QUESTIONS

The following picture emerges. Once one asks what types of representations are activated during linguistic processing, it becomes clear that one ought to distinguish (at least) between articulatory/motor, phonological, morpho-syntactic/lemma, and conceptual representations. The question of what conceptual representations are and how concepts are represented in the brain has garnered a lot of attention within philosophy and cognitive science and fuels the controversy between those who claim that conceptual representations are necessarily embodied and those who deny it. The cognitive potency and function of these other linguistic representations are less discussed in the literature.

Language unquestionably affords us cognitive benefits. Some of these benefits, we argue, are best explained on the assumption that language provides us with abstract and sparse representation. As outlined above, the availability and active usage of a linguistic code have been shown to enhance cognitive control. Plausible mechanisms of how language in general, and labels in particular, aid cognitive control are the increase of psychological distance by, arguably, activating abstract representations not immediately bound to action or perception. Those representations could encompass linguistic representations beyond conceptual ones, as, for example, abstract lexical (lemma) or phonological representations. Furthermore, the linguistic code is sparse and thus computationally cheap, yet powerful, as it exhibits combinatorial structure. Due to these properties, it may help to formulate, maintain, retrieve, and switch between task rules ("If stimulus X appears, then act in manner Y"). We conjecture that this is the basis of the cognitive functions of inner speech: based on the computational advantage of linguistic representations, inner speech enhances performance in problem-solving and other cognitively demanding tasks and augments cognitive control more generally.

The representational infrastructure of language, in overt or covert (inner) speech, consists of phonological, abstract-lexical, and syntactic representations, which may or may not be accompanied by embodied representations. The representations supporting cognitive control functions also seem to involve various kinds of abstraction. Assuming that representations with similar informational content are easier to map onto each other than to representations including different degrees of detail, it seems plausible that especially the more abstract properties of the linguistic code feed into the system guiding cognitive control.

All this is not to deny that detailed, embodied, sensory-motor representations may be of use, too. They may, occasionally, lead to deeper memory encoding, better retrieval, better multimodal processing, etc. Social cognition may also benefit from embodied linguistic representations as they may allow speakers to mentally align more easily by simulating similar experiences. They may also ease language acquisition and in many cases, language comprehension. Interpreting a novel metaphor or a poem, for example, may require that very rich, detailed, sensory-motor or affective representations are activated in order to understand the particular aspects of meaning that are targeted. For all that, the cognitive benefits of less embodied, abstract, and sparse representations are not to be denied either (Kompa, 2019). In the end, a more balanced and nuanced view that acknowledges that (i) multiple (types of) representations may be activated and drawn upon in a task-sensitive manner in linguistic processing and that (ii) there may be a gradient ranging from more embodied to more abstract (and maybe to different types of abstract) representations which all play (different) cognitive roles, may be the most promising route.

Also, note that we are not inferring the linguistic character of (some forms of) cognition from the experience of inner speech. It has been argued (Machery, 2005) that the phenomenology of inner speech does not provide evidence for the claim that cognition is linguistic, as the latter claim concerns the vehicles of thought (or the types of representations grounding conscious experience), which are not consciously accessibly. All we are claiming is that the findings of studies examining the cognitive benefits of inner speech seem to be best explained – at least as far as its effect on cognitive control is concerned – on the assumption that inner speech activates abstract linguistic representations of sorts. It is not an inference from phenomenology to neural implementation (that would indeed be invalid) but an inference to the best explanation of some of our cognitive accomplishments. Still, one might wonder whether those linguistic representations (allegedly) activated during control-demanding tasks (or in inner speech, for that matter) are consciously accessible. Now, while lemma or morpho-syntactic representations do not seem to reach the level of consciousness, phonological representations may

(but need not) do so. But then, there might be ways in which lemma representations (or some such thing) can be experienced, as suggested by studies on tip-of-the-tongue phenomena (cf. e.g., Vigliocco et al., 1997). This is slightly at odds with Carruther's claim that conscious access "always depends on attention directed at sensory representations of some sort" (Carruthers, 2018, p. 39). He argues that "most inner speech results from the mental rehearsal of speech actions" (Carruthers, 2018, p. 33), thus also involving the speech production system. In inner speech episodes, we activate but do not execute speech actions, and "[t]hese motor schemata are used to create a representation of what it would sound like if they were carried through to completion" (Carruthers, 2018, p. 33). On our view, executing may be stopped much earlier, maybe even before phonological representations become activated (thus at the level of lemma representations). Moreover, Carruthers has it that inner speech, being but a "copy of motor instructions" (Carruthers, 2018, p. 43), has no semantic content and needs to be interpreted by the speech comprehension system. This strikes us as a problematic idea, for why would one want to activate a speech action in inner speech that completely lacks content? What would be the point of that?

Finally, while the general conclusion that language aids cognitive control seems warranted, we also acknowledge that there are many open research questions with regard to how this comes about. For example, it is still not clear which properties of linguistic labels are responsible for their cognitive potency: is it their familiarity, their referential function, their phonological profile, their non-iconicity (i.e., the fact that they do not resemble what they denote), or something else still? Also, how do inner speech and outer speech relate to one another; what form can inner speech take and which purposes (over and above those indicated here) does it fulfill? How exactly are syntactic properties and combinatorial abilities implemented so as to mirror complex task structures? How exactly do lemma, conceptual, phonological, and other linguistic representations relate to one another with regard to their function for cognitive control? How do other cognitive domains, like memory, interact with language in the service of cognitive control? Does language play similar roles across different cognitive domains, e.g., cognitive control, memory, and learning? Future research will have to tackle these questions and will, hopefully, lead toward more detailed, explanatory models of how language and cognition interact.

AUTHOR CONTRIBUTIONS

The authors developed the ideas presented in the paper in close collaboration and as a result of intensive discussions on the topic. All authors contributed to the article and approved the submitted version.

FUNDING

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer GRK-2185/1 (DFG-Graduiertenkolleg Situated Cognition) [Funded by the German Research Foundation (DFG) – project number GRK-2185/1 (DFG Research Training Group Situated Cognition)].

ACKNOWLEDGMENTS

We are very grateful to the members of the RTG 'Situated cognition' at the universities of Osnabrueck and Bochum for helpful discussions. We would also like to thank the participants of the Workshop "The Cognitive Benefits of Language" which was hosted by Osnabrueck University in October 2019, for inspiring talks and debates; many thanks also to the two reviewers for valuable comments and feedback. Finally, we would like to thank Charles Lowe for carefully proofreading the manuscript.

REFERENCES

Addessi, E., Bellagamba, F., Delfino, A., De Petrillo, F., Focaroli, V., Macchitella, L., et al. (2014). Waiting by mistake: symbolic representation of rewards modulates intertemporal choice in capuchin monkeys, preschool children and adult humans. *Cognition* 130, 428–441. doi: 10.1016/j.cognition.2013.11.019

Alderson-Day, B., and Fernyhough, C. (2015). Inner speech: development, cognitive functions, phenomenology, and neurobiology. *Psychol. Bull.* 141, 931–965. doi: 10.1037/bul0000021

Althaus, N., and Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. *J. Exp. Child Psychol.* 151, 5–17. doi: 10.1016/j.jecp.2015.11.013

Anderson, M. (2014). *After phrenology: Neural reuse and the interactive brain*. Cambridge, MA: The MIT Press.

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266. doi: 10.1017/S0140525X10000853

Apperly, I. A., and Carroll, D. J. (2009). How do symbols affect 3- to 4-year-olds' executive function? Evidence from a reverse-contingency task. *Dev. Sci.* 12, 1070–1082. doi: 10.1111/j.1467-7687.2009.00856.x

Archibald, L. M. (2017). Working memory and language learning: a review. *Child Lang. Teach. Ther.* 33, 5–17. doi: 10.1177/0265659016654206

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* 16, 1818–1823. doi: 10.1016/j.cub.2006.07.060

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends Cogn. Sci.* 12, 193–200. doi: 10.1016/j.tics.2008.02.004

Badre, D., and D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J. Cogn. Neurosci.* 19, 2082–2099. doi: 10.1162/jocn.2007.19.12.2082

Badre, D., and Nee, D. E. (2018). Frontal cortex and the hierarchical control of behavior. *Trends Cogn. Sci.* 22, 170–188. doi: 10.1016/j.tics.2017.11.005

Baldo, J., Dronkers, N., Wilkins, D., Ludy, C., Raskin, P., and Kim, J. (2005). Is problem solving dependent on language? *Brain Lang.* 92, 240–250. doi: 10.1016/j.bandl.2004.06.103

Barkley, R. A. (2001). The executive functions and self-regulation: an evolutionary neuropsychological perspective. *Neuropsychol. Rev.* 11, 1–29. doi: 10.1023/A:1009085417776

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi: 10.1017/S0140525X99002149

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychon. Bull. Rev.* 23, 1122–1142. doi: 10.3758/s13423-016-1028-3

Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2012). "Language and simulation in conceptual processing" in *Symbols and embodiment: Debates on meaning and cognition*. eds. M. de. Vega, A. M. Glenberg and A. C. Graesser (Oxford: Oxford University Press), 245–283.

Barsalou, L. W., and Wiemer-Hastings, K. (2005). "Situating abstract concepts" in *Grounding cognition: The role of perception and action in memory, language, and thinking*. eds. D. Pecher and R. A. Zwaan (Cambridge: Cambridge University Press), 129–163.

Bedny, M., Koster-Hale, J., Elli, G., Yazzolino, L., and Saxe, R. (2019). There's more to "sparkle" than meets the eye: knowledge of vision and light verbs among congenitally blind and sighted individuals. *Cognition* 189, 105–115. doi: 10.1016/j.cognition.2019.03.017

Bermudez, J. L. (2018). "Inner speech, determinacy, and thinking consciously about thoughts" in *Inner speech: New voices*. eds. P. Langland-Hassan and A. Vicente (Oxford: Oxford University Press), 199–220.

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychon. Bull. Rev.* 23, 1096–1108. doi: 10.3758/s13423-015-0909-1

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., and Tummolini, L. (2019). Words as social tools: language, sociality and inner grounding in abstract concepts. *Phys. Life Rev.* 29, 120–153. doi: 10.1016/j.plrev.2018.12.001

Borghi, A. M., Barca, L., Binkofski, F., and Tummolini, L. (2018a). Abstract concepts, language and sociality: from acquisition to inner speech. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 373:20170134. doi: 10.1098/rstb.2017.0134

Borghi, A. M., Barca, L., Binkofski, F., and Tummolini, L. (2018b). Varieties of abstract concepts: development, use and representation in the brain. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 373:20170121. doi: 10.1098/rstb.2017.0121

Bourguignon, N. J., and Gracco, V. L. (2019). A dual architecture for the cognitive control of language: evidence from functional imaging and language production. *NeuroImage* 192, 26–37. doi: 10.1016/j.neuroimage.2019.02.043

Boysen, S. T., Berntson, G. G., Hannan, M. B., and Cacioppo, J. T. (1996). Quantity-based interference and symbolic representations in chimpanzees (*Pan troglodytes*). *J. Exp. Psychol. Anim. Behav. Process.* 22, 76–86. doi: 10.1037/0097-7403.22.1.76

Buckner, C. (2018). Empiricism without magic: transformational abstraction in deep convolutional neural networks. *Synthese* 195, 5339–5372. doi: 10.1007/s11229-018-01949-1

Caramazza, A., and Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon. *Cognition* 64, 309–343. doi: 10.1016/S0010-0277(97)00031-0

Carlson, S. M., and Beck, D. M. (2009). "Symbols as tools in the development of executive function" in *Private speech, executive functioning, and the development of verbal self-regulation*. eds. A. Winsler, C. Fernyhough and I. Montero (Cambridge: Cambridge University Press), 163–175.

Carlson, S. M., Davis, A. C., and Leach, J. G. (2005). Less is more: executive function and symbolic representation in preschool children. *Psychol. Sci.* 16, 609–616. doi: 10.1111/j.1467-9280.2005.01583.x

Carruthers, P. (2002). The cognitive functions of language. *Behav. Brain Sci.* 25, 657–674. doi: 10.1017/S0140525X02000122

Carruthers, P. (2018). "The causes and contents of inner speech" in *Inner Speech: New Voices*. eds. P. Langland-Hassan and A. Vicente (Oxford: Oxford University Press), 31–52.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton & Co.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

Christoff, K., and Gabrieli, J. D. E. (2000). The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology* 28, 168–186. doi: 10.3758/BF03331976

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends Cogn. Sci.* 10, 370–374. doi: 10.1016/j.tics.2006.06.012

Cohen, J. D. (2017). "Cognitive control" in *Wiley handbook of cognitive control*. ed. T. Egner (Chichester, UK: John Wiley & Sons, Ltd), 1–28.

Cole, M. W., Laurent, P., and Stocco, A. (2013). Rapid instructed task learning: a new window into the human brain's unique capacity for flexible cognitive control. *Cogn. Affect. Behav. Neurosci.* 13, 1–22. doi: 10.3758/s13415-012-0125-7

Connell, L. (2019). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Lang. Cogn. Neurosci.* 34, 1308–1318. doi: 10.1080/23273798.2018.1471512

Cragg, L., and Nation, K. (2010). Language and the development of cognitive control. *Top. Cogn. Sci.* 2, 631–642. doi: 10.1111/j.1756-8765.2009.01080.x

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* 93, 283–321. doi: 10.1037/0033-295X.93.3.283

Desai, R. H., Reilly, M., and Van Dam, W. (2018). The multifaceted abstract brain. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 373:20170122. doi: 10.1098/rstb.2017.0122

Diaz, R. M., and Berg, L. (1992). *Private speech: From social interaction to self-regulation*. Mahwah, NJ: Laurence Erlbaum Associates.

Doebel, S., Dickerson, J. P., Hoover, J. D., and Munakata, Y. (2018). Using language to get ready: familiar labels help children engage proactive control. *J. Exp. Child Psychol.* 166, 147–159. doi: 10.1016/j.jecp.2017.08.006

Dove, G. (2009). Beyond perceptual symbols: a call for representational pluralism. *Cognition* 110, 412–431. doi: 10.1016/j.cognition.2008.11.016

Dove, G. (2011). On the need for embodied and dis-embodied cognition. *Front. Psychol.* 1:242. doi: 10.3389/fpsyg.2010.00242

Dove, G. (2014). Thinking in words: language as an embodied medium of thought. *Top. Cogn. Sci.* 6, 371–389. doi: 10.1111/tops.12102

Dove, G. (2016). Three symbol ungrounding problems: abstract concepts and the future of embodied cognition. *Psychon. Bull. Rev.* 23, 1109–1121. doi: 10.3758/s13423-015-0825-4

Dove, G. (2018). Language as a disruptive technology: abstract concepts, embodiment and the flexible mind. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 373:20170135. doi: 10.1098/rstb.2017.0135

Dove, G. (2019). More than a scaffold: language is a neuroenhancement. *Cogn. Neuropsychol.* 1-24. doi: 10.1080/02643294.2019.1637338

Egner, T. (ed.) (2017). *The Wiley handbook of cognitive control*. Chichester. UK: John Wiley & Sons, Ltd.

Emerson, M. J., and Miyake, A. (2003). The role of inner speech in task switching: a dual-task investigation. *J. Mem. Lang.* 48, 148–168. doi: 10.1016/S0749-596X(02)00511-9

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Front. Psychol.* 5:335. doi: 10.3389/fpsyg.2014.00335

Fedorenko, E., and Varley, R. (2016). Language and thought are not the same thing: evidence from neuroimaging and neurological patients. *Ann. N. Y. Acad. Sci.* 1369, 132–153. doi: 10.1111/nyas.13046

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., et al. (2016). Concept representation reflects multimodal abstraction: a framework for embodied semantics. *Cereb. Cortex* 26, 2018–2034. doi: 10.1093/cercor/bhv020

Fernyhough, C. (2009). "Dialogic thinking" in *Private speech, executive functioning, and the development of verbal self-regulation*. eds. A. Winsler, C. Fernyhough and I. Montero (Cambridge: Cambridge University Press), 42–52.

Fernyhough, C. (2016). *The voices within*. London: Profile Books.

Gade, M., and Paelecke, M. (2019). Talking matters—evaluative and motivational inner speech use predicts performance in conflict tasks. *Sci. Rep.* 9:9531. doi: 10.1038/s41598-019-45836-2

Garrett, M. F. (1976). "Syntactic processes in sentence production" in *New approaches to language mechanisms*. eds. R. J. Wales and E. Walker (Amsterdam: North-Holland Publishing Company).

Gauker, C. (2018). "Inner speech as the internalization of outer speech" in *Inner speech: New voices*. eds. P. Langland-Hassan and A. Vicente (Oxford: Oxford University Press).

Hagoort, P. (2005). On broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423. doi: 10.1016/j.tics.2005.07.004

Hauk, O., Johnsrude, I., and Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307. doi: 10.1016/S0896-6273(03)00838-9

Huang, L., and Awh, E. (2018). Chunking in working memory via content-free labels. *Sci. Rep.* 8:23. doi: 10.1038/s41598-017-18157-5

Hurlburt, R. T., and Akhter, S. A. (2008). Unsymbolized thinking. *Conscious. Cogn.* 17, 1364–1374. doi: 10.1016/j.concog.2008.03.021

Hurlburt, R. T., Heavey, C. L., and Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Conscious. Cogn.* 22, 1477–1494. doi: 10.1016/j.concog.2013.10.003

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Res.* 1146, 2–22. doi: 10.1016/j.brainres.2006.08.111

Jackendoff, R. (2017). In defense of theory. *Cogn. Sci.* 41, 185–212. doi: 10.1111/cogs.12324

Jeon, H. -A., and Friederici, A. D. (2013). Two principles of organization in the prefrontal cortex are cognitive hierarchy and degree of automaticity. *Nat. Commun.* 4:2041. doi: 10.1038/ncomms3041

Jeon, H. -A., and Friederici, A. D. (2015). Degree of automaticity and the prefrontal cortex. *Trends Cogn. Sci.* 19, 244–250. doi: 10.1016/j.tics.2015.03.003

Jirak, D., Menz, M. M., Buccino, G., Borghi, A. M., and Binkofski, F. (2010). Grasping language—a short story on embodiment. *Conscious. Cogn.* 19, 711–720. doi: 10.1016/j.concog.2010.06.020

Kharitonova, M., Chien, S., Colunga, E., and Munakata, Y. (2009). More than a matter of getting 'unstuck': flexible thinkers use more abstract representations than perseverators. *Dev. Sci.* 12, 662–669. doi: 10.1111/j.1467-7687.2008.00799.x

Kharitonova, M., and Munakata, Y. (2011). The role of representations in executive function: investigating a developmental link between flexibility and abstraction. *Front. Psychol.* 2:347. doi: 10.3389/fpsyg.2011.00347

Kirkham, N. Z., Cruess, L., and Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Dev. Sci.* 6, 449–467. doi: 10.1111/1467-7687.00300

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science* 302, 1181–1185. doi: 10.1126/science.1088545

Kompa, N. A. (2019). Language and embodiment—or the cognitive benefits of abstract representations. *Mind Lang.* doi: 10.1111/mila.12266

Kray, J., Eber, J., and Karbach, J. (2008). Verbal self-instructions in task switching: a compensatory tool for action-control deficits in childhood and old age? *Dev. Sci.* 11, 223–236. doi: 10.1111/j.1467-7687.2008.00673.x

Kray, J., and Ferdinand, N. K. (2013). How to improve cognitive control in development during childhood: potentials and limits of cognitive interventions. *Child Dev. Perspect.* 7, 121–125. doi: 10.1111/cdep.12027

Kray, J., Kipp, K. H., and Karbach, J. (2009). The development of selective inhibitory control: the influence of verbal labeling. *Acta Psychol.* 130, 48–57. doi: 10.1016/j.actpsy.2008.10.006

Kray, J., Schmitt, H., Heintz, S., and Blaye, A. (2015). Does verbal labeling influence age differences in proactive and reactive cognitive control? *Dev. Psychol.* 51, 378–391. doi: 10.1037/a0038795

Kriete, T., Noelle, D. C., Cohen, J. D., and O'Reilly, R. C. (2013). Indirection and symbol-like processing in the prefrontal cortex and basal ganglia. *Proc. Natl. Acad. Sci.* 110, 16390–16395. doi: 10.1073/pnas.1303547110

Labuschagne, L. G., Cox, T. -J., Brown, K., and Scarf, D. (2017). Too cool? Symbolic but not iconic stimuli impair 4-year-old children's performance on the delay-of-gratification choice paradigm. *Behav. Process.* 135, 36–39. doi: 10.1016/j.beproc.2016.11.014

Lambon Ralph, M. A., Jefferies, E., Patterson, K., and Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55. doi: 10.1038/nrn.2016.150

LaTourrette, A., and Waxman, S. R. (2019). A little labeling goes a long way: semi-supervised learning in infancy. *Dev. Sci.* 22:e12736. doi: 10.1111/desc.12736

Laurent, L., Millot, J. -L., Andrieu, P., Camos, V., Floccia, C., and Mathy, F. (2016). Inner speech sustains predictable task switching: direct evidence in adults. *J. Cogn. Psychol.* 28, 585–592. doi: 10.1080/20445911.2016.1164173

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, Massachusetts: The MIT Press.

Locke, J. (1979). *An essay concerning human understanding*. ed. P. H. Nidditch (Oxford: Clarendon Press).

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Psychol.* 3:54. doi: 10.3389/fpsyg.2012.00054

Lupyan, G., and Bergen, B. (2016). How language programs the mind. *Top. Cogn. Sci.* 8, 408-424. doi: 10.1111/tops.12155

Lupyan, G., and Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 373:20170137. doi: 10.1098/rstb.2017.0137

Machery, E. (2005). You don't know how you think: introspection and language of thought. *Br. J. Philos. Sci.* 56, 469–485. doi: 10.1093/bjps/axi130

Machery, E. (2006). Two dogmas of neo-empiricism. *Philos. Compass* 1, 398–412. doi: 10.1111/j.1747-9991.2006.00030.x

Machery, E. (2007). Concept empiricism: a methodological critique. *Cognition* 104, 19–46. doi: 10.1016/j.cognition.2006.05.002

Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychon. Bull. Rev.* 23, 1090–1095. doi: 10.3758/s13423-015-0878-4

Mahon, B. Z. (2015). What is embodied about cognition? *Lang. Cogn. Neurosci.* 30, 420–429. doi: 10.1080/23273798.2014.987791

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70. doi: 10.1016/j.jphysparis.2008.03.004

Mahon, B. Z., and Hickok, G. (2016). Arguments about the nature of concepts: symbols, embodiment, and beyond. *Psychon. Bull. Rev.* 23, 941–958. doi: 10.3758/s13423-016-1045-2

Massaro, D. W., and Cohen, M. M. (1983). Phonological context in speech perception. *Percept. Psychophys.* 34, 338–348. doi: 10.3758/BF03203046

Meteyard, L., Cuadrado, S. R., Bahrami, B., and Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48, 788–804. doi: 10.1016/j.cortex.2010.11.002

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Miller, S. E., and Marcovitch, S. (2011). Toddlers benefit from labeling on an executive function search task. *J. Exp. Child Psychol.* 108, 580–592. doi: 10.1016/j.jecp.2010.10.008

Miozzo, M., and Caramazza, A. (1997). Retrieval of lexical—syntactic features in tip-of-the tongue states. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 1410–1423. doi: 10.1037/0278-7393.23.6.1410

Mischel, W., Shoda, Y., and Rodriguez, M. (1989). Delay of gratification in children. *Science* 244, 933–938. doi: 10.1126/science.2658056

Miyake, A., Emerson, M. J., Padilla, F., and Ahn, J. (2004). Inner speech as a retrieval aid for task goals: the effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychol.* 115, 123–142. doi: 10.1016/j.actpsy.2003.12.004

Miyake, A., and Friedman, N. P. (2012). The nature and organization of individual differences in executive functions. *Curr. Dir. Psychol. Sci.* 21, 8–14. doi: 10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., and Fujimura, O. (1975). An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. *Percept. Psychophys.* 18, 331–340. doi: 10.3758/BF03211209

Müller, U., Zelazo, P. D., Lurye, L. E., and Liebermann, D. P. (2008). The effect of labeling on preschool children's performance in the dimensional change card sort task. *Cogn. Dev.* 23, 395–408. doi: 10.1016/j.cogdev.2008.06.001

Noë, A., and Thompson, E. (2004). Are there neural correlates of consciousness? *J. Conscious. Stud.* 11, 3–28.

Pauls, L. J., and Archibald, L. M. D. (2016). Executive functions in children with specific language impairment: a meta-analysis. *J. Speech Lang. Hear. Res.* 59, 1074–1086. doi: 10.1044/2016_JSLHR-L-15-0174

Perler, D., and Haag, J. (eds.) (2010). *Ideen–Representationalismus in der Frühen Neuzeit*. Berlin/New York: De Gruyter.

Pinker, S. (1991). Rules of language. *Science* 253, 530–535. doi: 10.1126/science.1857983

Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge, MA: The MIT Press.

Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582. doi: 10.1038/nrn1706

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn. Sci.* 17, 458–470. doi: 10.1016/j.tics.2013.06.004

Roelofs, A., Meyer, A. S., and Levelt, W. J. M. (1998). A case for the lemma/lexeme distinction in models of speaking: comment on Caramazza and Miozzo (1997). *Cognition* 69, 219–230. doi: 10.1016/s0010-0277(98)00056-0

Rouault, M., and Koechlin, E. (2018). Prefrontal function and cognitive control: from action to language. *Curr. Opin. Behav. Sci.* 21, 106–111. doi: 10.1016/j.cobeha.2018.03.008

Rougier, N. P., Noelle, D. C., Braver, T. S., Cohen, J. D., and O'Reilly, R. C. (2005). Prefrontal cortex and flexible cognitive control: rules without symbols. *Proc. Natl. Acad. Sci.* 102, 7338–7343. doi: 10.1073/pnas.0502455102

Saeki, E., Baddeley, A. D., Hitch, G. J., and Saito, S. (2013). Breaking a habit: a further role of the phonological loop in action control. *Mem. Cogn.* 41, 1065–1078. doi: 10.3758/s13421-013-0320-y

Sanford, A. J. (2008). "Defining embodiment in understanding" in *Symbold and embodiment: Debates on meaning and cognition*. eds. M. de. Vega, A. Glenberg and A. Graesser (Oxford: Oxford University Press), 181–194.

Saysani, A., Corballis, M. C., and Corballis, P. M. (2018). Colour envisioned: concepts of colour in the blind and sighted. *Vis. Cogn.* 26, 382–392. doi: 10.1080/13506285.2018.1465148

Schyns, P. G., Goldstone, R. L., and Thibaut, J. P. (1998). The development of features in object concepts. *Behav. Brain Sci.* 21, 1–17. doi: 10.1017/s0140525x98000107

Sulpizio, S., Del Maschio, N., Fedeli, D., and Abutalebi, J. (2020). Bilingual language processing: a meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 108, 834–853. doi: 10.1016/j.neubiorev.2019.12.014

van't Wout, F., and Jarrold, C. (2020). The role of language in novel task learning. *Cognition* 194:104036. doi: 10.1016/j.cognition.2019.104036

Vigliocco, G., Antonini, T., and Garrett, M. F. (1997). Gramatical gender is on the tip of Italian tongues. *Psychol. Sci.* 8, 314–317.

Vigliocco, G., Vinson, D. P., Lewis, W., and Garrett, M. F. (2004). Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. *Cogn. Psychol.* 48, 422–488. doi: 10.1016/j.cogpsych.2003.09.001

Vygotski, L. S. (1986). *Thought and language*. Cambridge, Massachusetts: The MIT Press.

Werker, J. F., and Tees, R. C. (1984). Cross-language speech-perception—evidence for perceptual reorganization during the 1st year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3

Wiley, N. (2016). *Inner speech and the dialogical self*. Philadelphia: Temple University Press.

Winsler, A., Fernyhough, C., and Montero, I. (2009). *Private speech, executive functioning, and the development of verbal self-regulation*. Cambridge: Cambridge University Press.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition* 85, 223–250. doi: 10.1016/S0010-0277(02)00109-9

Yee, E., and Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychon. Bull. Rev.* 23, 1015–1027. doi: 10.3758/s13423-015-0948-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 21 August 2020
doi: 10.3389/fpsyg.2020.02140

Meshed Architecture of Performance as a Model of Situated Cognition

Shaun Gallagher[1,2]* *and Somogy Varga*[3]

[1] Philosophy, University of Memphis, Memphis, TN, United States, [2] School of Liberal Arts, University of Wollongong, Wollongong, NSW, Australia, [3] Department of Philosophy and the History of Ideas, Aarhus University, Aarhus, Denmark

In this paper, we engage in a reciprocal analysis of situated cognition and the notion of "meshed architecture" as found in performance studies (Christensen et al., 2016). We start with an account of various conceptions of situated cognition using the distinction between functional integration, which characterizes how an agent dynamically organizes to couple with its environment, and task dependency, which specifies various constraints and structures imposed by the environment (see Slors, 2019). We then exploit the concept of a meshed architecture as a model that provides a more focused analysis of situated cognition and performance. Through this analysis, we show how the model of meshed architecture can be enhanced through (1) the involvement of a more complex set of cognitive processes, (2) a form of intrinsic control, (3) the influence of affective factors, and (4) the role of factors external to the performer. The aim of this paper, then, is twofold: first to work out an enhanced conception of the model of meshed architecture by taking into consideration a number of factors that clarify its situated nature, and second, to use this model to provide a richer and more definitive understanding of the meaning of situated cognition. Thus, we argue that this reciprocal analysis gives us a very productive way to think about how various elements come together in skilled action and performance but also a detailed way to characterize situated cognition.

Keywords: situated cognition, performance, task dependency, body schema, functional integration, meshed architecture

OPEN ACCESS

Edited by:
Achim Stephan,
University of Osnabrück, Germany

Reviewed by:
Matthew Ratcliffe,
University of York, United Kingdom
Wendy Wilutzky,
University of Bergen, Norway

***Correspondence:**
Shaun Gallagher
s.gallagher@memphis.edu

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *08 May 2020*
Accepted: *31 July 2020*
Published: *21 August 2020*

Citation:
Gallagher S and Varga S (2020) Meshed Architecture of Performance as a Model of Situated Cognition. Front. Psychol. 11:2140. doi: 10.3389/fpsyg.2020.02140

WHAT'S THE SITUATION?

Embodied, *embedded*, *extended*, and *enactive* approaches to cognition comprise a loose-knit group of research endeavors that endorse the view that the organism's body and parts of its environment actively participate in the execution of cognition. They differ in their views about how mind and world are entangled. For example, some endorse *epistemological inseparability* (i.e., no full understanding of cognitive processes is possible by studying exclusively what is occurring inside the head) while others also endorse *ontological inseparability* (i.e., the realizers of cognitive processes can sometimes include parts of the body and the environment) (Varga, 2019). Most would agree that such approaches can be grouped along these two claims, but some have argued that the distinctions

in the literature are muddled (Rowlands, 2010). For example, some maintain that work in situated cognition investigates cognitive extensions (Clark and Wilson, 2009) while others consider extended cognition as a distinct class of situated cognition (Robbins and Aydede, 2009, p. 3).

Situated cognition, however, can be considered a broad umbrella term that covers all of these various approaches. As such, it is multifaceted. We might think of it in terms of how environmental features both constrain and enable our cognitive processes. On the constraint side, we can think of various material and structural features as directing us to a specific set of affordances, not only for our perceptions and actions, but also for our deliberations and imaginings. At the same time, these same affordances are enabling of our actions and cognitive activities. It is possible to think of these relations in terms of extended or distributed cognition. Various instruments allow us to engage in epistemic actions (Kirsh and Maglio, 1994), and for some cognitive tasks, we require the use of such instruments. To do the math, we may need paper, pencil, abacus, or some form of machine. To solve a problem, we may rely not only on such tools but also on other people or team members with whom to interact, as well as on normative practices and institutions (understood as cognitive institutions – see Gallagher, 2013; Slaby and Gallagher, 2015). At the same time, these practices and institutions may define specific tasks and place limitations on how we approach a problem or on our style of problem solving.

Thinking of situated cognition in this way, we can define our cognitive engagements as spanning a range between *functional integration* and *task dependency*. Marc Slors (2019) has recently clarified these concepts in his analysis of cognitive institutions. We think they can generalize to situated cognition more broadly. Following Slors, for example, we can distinguish between (1) the extended mind approach which starts from the single agent and explains how institutions extend the agent's cognition (Clark and Chalmers, 1998) and (2) the distributed cognition or systems-based approach that shows how cognitive systems emerge from the integration of individual agents (Hutchins, 2014). In this context, Slors defines functional integration as "the extent to which the execution of tasks involves coupling with items external to the brain and body" (Slors, 2019, p. 1189). A high degree of functional integration means that the cognitive process is constituted by this coupling such that without the external resource, we would be unable to engage in the particular activity, while low functional integration signifies an *enabling* relation such that the external resource simply facilitates our activity. In contrast, task dependency

> is the extent to which the intelligibility of a task depends on a larger whole of coordinated tasks. Task dependency is a notion that is connected with coordination and planning. It is a normative notion in the sense that high task dependency means that tasks play specific roles in the overall organization of a cognitive system or a cultural cognitive ecosystem, roles that can be played properly or improperly (Slors, 2019, p. 1190).

The legal system, for example, understood as a cognitive institution (Gallagher, 2013) is characterized by high task dependency. Accordingly, to understand what an attorney does requires an understanding of how that role is linked to the roles played by other people, such as judges and clerks, as well as to a codified body of laws and customs. What one might accomplish in this system will depend upon the structure of the particular situation that constitutes a social-normative or institutional practice.

Situated cognition, then, can be categorized by varying degrees of functional integration and task dependency (**Table 1**).

Embedded cognition is defined by a low functional integration with various resources that nonetheless enables the performance of cognitive activities and where such activities are intelligible without reference to the institutional structure (low task dependency). Distributed cognition, in contrast, involves the right coupling between distributed components, such as artifacts or other agents in a highly functionally integrated system that also requires high task dependency such that the action of any one individual cannot be understood without reference to others' activities. Extended cognition (in Clark and Chalmers' sense) involves high functional integration. Otto, for example, is tightly coupled with his notebook, which allows an extension of his cognitive processes, even if writing and consulting a notebook are not processes that necessarily depend on the roles or tasks of others to be intelligible. "Symbiotic cognition," as Slors terms it, is found in cognitive institutions. In symbiotic cognition, characterized by high task dependency, an individual's cognitive processes acquire meaning only in a matrix of interrelationships with the activities of others but do not require a high degree of functional integration.

> Every participant in a symbiotic system profits from whatever the system as a whole offers (e.g., education, justice, social coordination) while contributing only a small part. The tasks, jobs and roles of others in the system co-define and enable one's own task, but one does *not* have to perform them or even think about them, while nevertheless benefiting from the overall outcome of the system (Slors, 2019, p. 1198).

Although this is a productive analysis, it is an oversimplification to think of cognitive institutions as strictly symbiotic or characterized specifically in terms of high task dependency (see Gallagher et al., 2019; Petracca and Gallagher, 2020). As Slors (2019, p. 1190) rightly indicates, "both functional integration and task-dependency come in degrees," and it seems right to think that a cognitive institution, or situated cognition more generally, will always involve varying degrees of task dependency and functional integration (also see Slors, 2020). Furthermore, how one understands a system will depend on where in the system one is looking or perhaps from what epistemic perspective one is looking. Specifically, the distinction between agent-centered and systems-based perspectives involves

TABLE 1 | Forms of situated cognition (from Slors, 2019, p. 1191).

	Low functional integration	**High functional integration**
Low task dependency	Embedded cognition	Extended cognition
High task dependency	Symbiotic cognition	Distributed cognition

different epistemological perspectives that may serve different research agendas but does not necessarily define particular institutional processes. From a systems perspective, a system may involve high task dependency. But from an agent-centered perspective, one may see a significant degree of functional integration that defines that agent's work.

To motivate our strategy of looking at performance studies to provide some detail in this regard, consider that the notion of task dependency, where action may be defined by a particular role in the performance, is clearly relevant to different types of performance. Of course, task dependency will vary across different types of performance, but different tasks or roles, performed by specific participants, may still be clearly defined, for example, when one is playing football, dancing or acting on stage, or playing a concert in the Sydney Opera House or in the local pub. Specifically, one will find variations in the proportion of functional integration versus task dependency. In a standard tonal jazz performance, for example, task dependency may take the lead. There is a structure to the performance – first, playing the "head," a statement of the main melodic line; then solos where each performer takes turns following rules concerning harmonic and relatively consistent chord changes; and then, after each performer has taken one or two choruses, the group plays the outro, to end the piece. If one is performing a solo improvisation, without an ensemble, team, or musical group, then task dependency may approach zero, and functional integration may be everything. The latter is a different kind of situation from performing with others; but in each case, the performance and the cognition that goes with it are situated within some variable proportion of functional integration and task dependency.

Functional integration defines how individual agents engage with the various elements of the system and, in so doing, enact the system, which loops back to define performance in specific tasks. An explanation that simply highlights the distinction between functional integration and task dependency, however, remains a somewhat abstract account of situated cognition and is not sufficient to account for how situated agents actually couple to environments or how tasks that are institutionally or more broadly socially, culturally, or normatively defined actually shape that environment. At best, it is an initial specification that requires a more detailed account of how it all works. That is, even if an analysis in terms of functional integration and task dependency provides a productive way to categorize different conceptions of situated cognition, it does not explain precisely how an agent functionally couples with the environment or enters into task-related processes. For example, an adequate concept of functional integration needs to include more than just an account of organism–environment coupling; it also needs to explain how the agentive organism itself is integrated so as to facilitate this coupling.

In the remainder of the paper, we want to provide an account of what it is to be a situated agent engaged in some task or performance that involves varying degrees of task dependency and functional integration. To do this, we turn to the model of a meshed architecture developed in performance studies. We propose that by going into some detail about this model, we can flesh out some of the important aspects of situated cognition. In this respect, we argue that there can be a reciprocal or mutual enlightenment between studies of performance and the theory of situated cognition.

MESHED ARCHITECTURE IN PERFORMANCE

As long as an agent is not simply a cog in the machine (an indifferent functional part of the system), one can think of her as a skilled performer or as someone who practices some degree of skilled activity. As we will see, this is one way to characterize functional integration, and it involves something more than simply fulfilling a predefined task in an automatic way, although from a systems perspective, this may sometimes appear to be what is happening. In this respect, we want to rule out the idea of a zero-intelligence agent (see Gode and Sunder, 1993) – that is, an agent who, to perform a task, acts in a purely automatic way and whose performance would involve no cognitive contribution. Functional integration is something more than this and involves a process that is both more complex and more subtle. To make this clear, we turn to a debate in performance theory and the specific model of a meshed architecture to clarify the perspective of a situated agent.

In the area of performance studies, Hubert Dreyfus' well-known analysis of expertise would come close to the zero-intelligence agent. Dreyfus argued that expert performance involves being mindlessly in the flow, since any form of reflective cognition would be disruptive of performance. He regards subjectivity as a lingering ghost of the mental and denies that there is any awareness in absorbed coping (Dreyfus, 2007, p. 373). On this model, as long as things go smoothly, the agent can be on automatic pilot; there is no need for self-consciousness – the latter is called into action only when the agent detects something going wrong (Dreyfus, 2007, p. 377; see Dreyfus, 2005). At the extreme, this view suggests that expert performance is simply a mindless being in the flow. The elite Sri Lankan cricketer Kumar Sangakkara expresses this view: "Basically in batting, you have to be mindless. You've done all the practice, you have your muscle memory and your reflexes are more than quick to deal with any kind of delivery. You've got to let your body do all those things by itself without letting your mind take control" (Sadikot, 2014).

In contrast, empirical and phenomenological studies of athletics, dance, theater, and musical performance suggest that performance is not mindless; there is always a cognitive element in performance. Moreover, the cognition involved is always a situated cognition. For example, John Sutton et al. (2011) develop a mindful conception of expert skilled performance. It is not just trained habit that allows an expert player of cricket or baseball to hit a hard fastball (which may be traveling at 140 km/h). In order to hit the ball with precision to a particular location, the batter must draw on current context and the conditions that are relevant to the game. Her performance is "fast enough to be a reflex, yet it is perfectly context-sensitive. This kind of context sensitivity requires some forms of mindedness – [an] interpenetration of

thought and action exemplified in open skills" (Sutton et al., 2011, p. 80). The expert batter cannot be on automatic pilot; being on automatic pilot would reduce functional integration to being just one piece of machinery fit to task. Batting skill within the context of a game, for example, involves some mindful strategic sense of where the batter will hit the ball in any particular instance.

> Skill is not a matter of bypassing explicit thought, to let habitual actions run entirely on their own, but of building and accessing flexible links between knowing and doing. The forms of thinking and remembering which can, in some circumstances, reach in to animate the subtle kinesthetic mechanisms of skilled performance must themselves be redescribed as active and dynamic (Sutton et al., 2011, p. 95).

Automaticity, therefore, cannot address variability or differences in the agentive situation. Skill and innovative performance require flexibility – the expert batter is aware of the specifics of the situation (including his own skills) and is capable of on-the-fly, explicit, considered awareness which allows for strategic decision making in the flow of performance. This includes elective "target control for some features, such as goal, one or more parameters of execution, like timing, force, a variation in the sequence, and so on" (Christensen et al., 2016, p. 50). In this respect, "expert performers precisely counteract automaticity, because it limits their ability to make specific adjustments on the fly. . .. Just because skillful action is usually pre-reflective, it does not have to be mindless" (Sutton et al., 2011, p. 95).

To say that functional integration is something more than automaticity in the context of skilled performance, then, motivates several questions. First, what are the cognitive processes involved, and second, how precisely do they "reach in" to the basic body-schematic processes of skilled performance?

Christensen et al. (2016) offer a helpful answer to the second question, developing the concept of a *meshed architecture* to explain the integration of perceptual and cognitive elements with body-schematic motoric processes. On this view, performance is neither fully automatic nor fully cognitive. They develop a hybrid view according to which "cognitive control reduces during skill learning as automatic control comes to play an increasing role, but cognitive control continues to make a substantial positive contribution at advanced levels of skill" (Christensen et al., 2016, p. 41). They propose a *meshed* functioning which involves "a broadly hierarchical division of control responsibilities" along a vertical axis, with top-down cognitive control "focused on strategic aspects of performance and [bottom-up] automatic processes more concerned with implementation" (Christensen et al., 2016, p. 43). Control is mediated, not by explicit inferences, but by "situated awareness," an awareness that is "constructed" over time with the help of attentional control.

To help us understand how the notion of a meshed architecture can generalize more broadly to situated cognition and contribute to an explanation of functional integration and task dependency discussed above, we propose the following clarifications. First, we suggest that the cognitive processes involved in performance are complex and varied and can include a full register that goes from explicit conscious control to implicit pre-reflective consciousness. Second, we argue that control is not just top-down but can also be intrinsic to bottom-up processes. Third, we argue for the importance of affective factors in modulating intrinsic control features. And fourth, especially in regard to situated cognition, it is even more important to consider that the mesh is complicated by a form of horizontal integration. The horizontal axis of integration includes ecological, social, and cultural/normative factors that extend beyond the performing agent but nonetheless constrain or contribute to performance. By making these clarifications, we hope to provide a more adequate view of how functional integration and task dependency work in situated cognition.

COMPLEX COGNITION

The notion of a meshed architecture has been applied to various types of performance, from athletic performance to the performing arts of acting, dance, and music. Different interpretations of a meshed architecture are possible, however, depending on how we answer the first question about how to understand the cognitive processes involved. Some theorists think of these processes in terms of high-order cognition. For example, in his discussion of theatrical acting, Cohen (2013, p. 33) refers to the actor's "preparatory thinking as she readies herself for the role, and in-performance thinking, which, in an ideal situation, is 'aligned' with the [performer's] action." For Cohen, when the actor's thinking is "properly aligned, her tasks are integrated" (Cohen, 2013, p. 16). This, as Tribble (2016) indicates in her discussion of the meshed architecture, would be a top-down process for Cohen, where low-order processes of embodied coping are modulated by higher-order, reflective cognitive aspects.

Likewise, Montero (2010, 2015) challenges the idea that expert performance is somehow effortless or thoughtless. She argues, in opposition to Dreyfus, that for expert dancers, reflection and body awareness are typically not detrimental to the performance. For Montero (2016, p. 38), optimal performance often coincides with reflective, thoughtful performance, where thoughtful means "self-reflective thinking, planning, predicting, deliberation, attention to or monitoring of . . . actions, conceptualizing . . . actions, control, trying, effort, having a sense of the self, and acting for a reason." Montero (2015, p. 90) pointed to qualitative studies in athletics where a more detailed type of conscious monitoring improves performance (also see Shusterman, 2008 for a similar account).

One could think of this as a type of vertical alignment between higher-order cognitive processes and lower-order motor control processes, with different degrees of integration between the higher- and lower-order processes. This is similar to what Christensen et al. (2016, p. 43) have in mind as they describe the mesh as a combination of cognitive (control-related) and automatic processes: thus, "controlled and automatic processes are closely integrated in skilled action, and . . . cognitive control directly influences motor execution in many cases." This divides the vertical into two poles: cognitive at the top, descending to do its job in a "smooth," "adaptive," or "effortful"

fashion (Christensen et al., 2016, p. 52), and automatic bodily processes at the bottom.

It is possible, however, that there are different degrees of vertical integration in the meshed architecture. Again, this goes back to how one answers the first question about the nature of the cognitive processes involved. The answer shifts between a phenomenology that involves a reflective monitoring and one that involves a more minimal pre-reflective awareness. For phenomenologists, pre-reflective self-awareness does not take the body as an intentional object; it rather involves a "performative awareness . . . that provides a sense that one is moving or doing something, not in terms that are explicitly about body parts, but in terms closer to the goal of the action" (Gallagher, 2005, p. 73). Legrand (2007, p. 512) described this self-awareness in the context of dance performance: "while dancing [a dancer] is intensively attending to [his body]. But he is not attending to it reflectively as an object. Rather, his [prereflective] awareness of his body as subject is heightened" (see Legrand and Ravn, 2009). The expert dancer keeps this awareness "at the front" of his experience without turning his action or his body into an explicit intentional object (Legrand, 2007, p. 512).

In these various accounts, it seems that what Christensen et al. (2016) call situated awareness can be a matter of degree, ranging from thoughtful, reflective consciousness to a thin performative pre-reflective awareness, with different gradations in between, allowing for such variations as selective target control, conscious monitoring, a sense of one's rightly configured body, performative awareness, and pre-reflective awareness. The phenomenology of performance may thus be complex and varied. Performers are able to shift across a full register, from explicit conscious control to implicit pre-reflective consciousness and to spontaneous body-schematic processes, adjusting their attunement to changing conditions through improvisation.

INTRINSIC CONTROL

One important question for clarifying the notion of functional integration, as we indicated above, is whether we should consider body-schematic processes as fully automatic. Christensen et al. (2016) mentioned this issue with reference to Fitts and Posner (1967, p. 14) who thought that component processes may automate and Jonides et al. (1985) and Logan (1985) who argued that motor control processes overall do not automate. Christensen, Sutton, and McIlwain seemed to treat body-schematic processes as fully automatic and therefore in need of cognitive control in the performance situation (see Stanley, 2011 for a similar view).

Evidence from kinematics, however, suggests that body-schematic processes are not fully automatic and instead are situation specific, adaptive, and highly dynamical, which facilitates movement in particular situations and for specific intentions. A particular action intention or goal requires the alignment of lots of moving parts in a controlled integration, across varying timescales, many of which are too fast for conscious control. In this respect, however, body-schematic processes are neither fully automatic (blindly doing the same thing in each circumstance and therefore requiring propositional guidance) nor "perfectly general" (Stanley, 2011) but rather include a specificity that depends on an "enormous number (which often reaches three figures) of degrees of freedom" (Bernstein, 1984), as well as a complex temporal organization involving anticipatory processes across skeletal geometry, kinematic phase constraints, muscular geometry, and the dynamics that characterize the relationship between kinematics and geometry (Berthoz, 2000; Gallagher and Aguda, 2020). These complex processes come to align with a particular intention, *not automatically* but in heedful attunement with the particularities of the situation.

Functional integration within such constraints may tune motoric organization to the point where it can become habitual – which may mean *close to* automatic, or automatic in some aspects, but not fully automatic. Merleau-Ponty (2012, p. 143) argued that a habit is formed when the body "acquires the power of responding with a certain type of solution to a certain form of situation." Habit involves an intelligent response, characterized by openness and adaptivity, so that in familiar or unfamiliar situations, the body learns to cope. As such, intelligence is built into the movement. Instead of blind automatic repetition, habit is intrinsically intelligent. John Dewey likewise distinguished between intelligent and routine habit.

> Repetition [i.e., automaticity] is in no sense the essence of habit. . . . The essence of habit is an acquired predisposition to *ways* or modes of response. . .. Habit means special sensitiveness or accessibility to certain classes of stimuli, standing predilections and aversions, rather than bare recurrence of specific acts (Dewey, 1922, p. 42).

On this view, performance involves not simply a top-down integration of cognition constraining or guiding automatic processes. Motoric processes in expert performance are already context sensitive, anchored in the situation, but at the same time smart, open, and adaptive, such that they elicit or shape or enable the cognitive elements required for performance. Not only are such cognitive elements, as already noted, complex, including heedful and goal-oriented forms of (attentive, perceptual) consciousness, selective target control, conscious monitoring of action, a sense of one's rightly configured body, and/or a heightened pre-reflective awareness, but also in such cases, mindfulness is not simply imported from the top; it is already built into the bottom, and again in some cases, such habitual processes may be what guides any need for more reflective cognitive processes. We can call this a kind of intrinsic control.

To summarize, for Christensen et al. (2016), the meshing involves a vertical axis of top-down cognitive control that introduces guidance for bodily processes that remain, at the bottom, automatic. This particular conception of the hybrid mesh, as Høffding and Satne (2019) suggested, is similar to the hybrid car that combines two different elements, battery and fuel. In contrast, they suggested that the mesh may be closer to a fusion – more like an okapi (a unique animal born of zebra and giraffe) than a hybrid car that alternates

between the current of automaticity and the fuel of high-octane cognition[1]. An okapi-style mesh, on our view, has a more integrated structure. Practiced and habitual movements (which are neither straightforwardly nor fully nor necessarily automatic) play an important role in an intrinsic control process. Variations in heedful and targeted (attentive, perceptual) awareness are constrained and enabled by a consolidation of fine, detailed motor control (body-schematic) processes, which are not perfectly general or automatic but attuned to the specifics of the situation.

AFFECT AND HORIZONTAL MESHING

We can get a better idea of what other factors contribute to the meshed architecture by considering an example of musical performance. Simon Høffding's (2019) study of the Danish String Quartet provided some evidence that the meshed architecture involves both a complex vertical and horizontal integration. Thus, for example, concerning the vertical axis, we find considerations, similar to those above, about the role of thoughtful performance ranging from explicit reflective thinking to pre-reflective awareness and, in some cases, a form of deep absorption where close to automatic processes of the body schema do most of the work. Along this line, Høffding and Satne (2019) interpreted the notion of a meshed architecture as focused on mediating processes rather than the all-or-nothing "automatic" versus "full cognitive" control (also see Salice et al., 2017).

The other factors that Høffding's analysis considers, in addition to the reciprocal vertical integration of cognition and body-schematic attunement, include affect but also the music itself and intersubjectivity, i.e., the other players. We conceive of the latter two factors as clearly on a horizontal axis which reaches aspects that most theorists would consider as constitutive of the agent's situation. Affectivity, however, is central and may define the vertical–horizontal intersection.

Affect in the broadest sense includes emotion processes but also more general and basic bodily states such as hunger, fatigue, and pain. Affect, or what Michelle Maiese (2018) calls "affective framing," shapes our ability to cope with the surrounding world (Ratcliffe, 2012; Colombetti, 2014) and, along with skills and habits, introduces possible modulations of functional integration with that world. Affect may work differently in different types of skilled actions, for example, in various athletic performances and in the different performing arts. The important differences may have to do with the way that affective factors are integrated with motoric/agentive factors – the kinetic and kinesthetic feelings associated with body-schematic processes and how all of these processes functionally integrate with environmental constraints and affordances. Affect/emotion may involve expressive movement, as in dance – movement that is like gesture and language but nonetheless depends on motor control – although it goes beyond simple motor control or instrumental action. There are different mixes or integrations of expressive and instrumental movements in the different contexts of performance – in athletics, dance, or musical performance, for example.

The body schema does not work independently to deliver technically proficient movement, to which an expressive style is then added as something motivated by specific and perhaps occasion-relative emotions. Affective processes directly shape body-schematic processes, slowing them or speeding them or leading them to a certain initial posture that may influence performance or change how agents are functionally situated. Accordingly, affect modulates functional integration. Affect and body-schematic processes are part of the vertical mesh in expert performance – but they also allow for an integration attuned to targets and environmental features in the performance situation.

In the context of musical performance, once we start to think about the music itself and the other performers, for example, we come to an enriched conception of the meshed architecture that incorporates a form of horizontal integration. In this respect, ecological, normative, cultural, and intersubjective aspects of the physical and social environments, including physical and social affordances, play a role and contribute to task-dependent structures in performance. For example, in Høffding's analysis, the musical instruments, the performance space, and the music itself shape the musical performance. The style of music, whether one is playing from a score, or whether improvisation occurs – these factors establish different roles and tasks and specify different possible dynamics in performance. All of this, in line with embodied-enactive conceptions of affordance-based action and cognition, as well as ecological psychology's conception of resonance, helps to show that what makes performance what it *is* is not entirely inside the performer, whether she be musician, dancer, athlete, or expert in everyday affairs. For example, the individual performer affectively resonates with and through the music. Playing the musical notes initiates a resonance between the sounds one creates and the musical sounds in the environment made by other musicians.

> This resonance may be driven by (1) consciously anticipated, and sometimes planned, notes and/or (2) feedback from awareness of the sounds that are actually created during performance. On one hand, as the music unfolds, the performance environment is constituted as a niche of musical affordances. The sounds that a musician produces could thus successfully or unsuccessfully resonate with the affordances in the environment. On the other hand, anticipatory processes and any short-term planning involved while playing suggest intraorganism resonant loops constantly underlying the performance (Ryan and Gallagher, 2020).

The combination of these respective elements is the mesh between anticipatory control, practiced/skilled bodily movements, and the affordances presented by the music and the environment more generally.

[1] Christensen and Sutton (2019) seemed to move closer to the okapi model. They relaxed the strong dualism between cognition and automaticity (an either–or arrangement), opting for more hybrid or pluralist (both–and) arrangements: "in which there are multiple levels of control, including lower-level, fast perception-action loops and higher-level loops that integrate more widely and process more abstract information, with the loops functioning in intimate interaction" (Christensen and Sutton, 2019, p. 160).

As one engages in a particular performance, one's agency (or sense of agency) may be modulated (causally influenced) by affect but also by the quality and quantity of affordances available. When, for example, the performer "can 'feel' that her motor system has the right configuration" (Christensen et al., 2016), this configuration is just the right one to mesh with the specifics of the performer's physical and social environments. Høffding (2019, p. 244) called this "interkinaesthetic affectivity" (see Salice et al., 2017; Høffding and Satne, 2019). Neither body-schematic processes nor affective processes are isolated from the agent's environment; rather, they are attuned to both stabilities and variations in environmental factors, including other agents. The performance and the cognition involved in it are situated, i.e., functionally integrated with the environment. Likewise, environmental factors, including music and interpersonal relations, can facilitate emotion regulation or regulation of affect more generally (Krueger, 2014, 2019).

The environment where performance takes place is not only physically but also socially, culturally, and normatively defined. Performance in a concert hall or in a church may be quite different from performance in a stadium or a pub or in the open air. That we are playing music with others, and who those others are, how skilled they are, and how long we have interacted with them – all of these factors can impact performance (Clarke et al., 2015). If one is playing music with others, there will be an intersubjective and affective resonance between an individual's performance and the performance of other musicians. This may be mediated by the music itself, by conscious, non-conscious, and/or non-verbal perceptual cues in the others' embodied performance (see Høffding, 2019; Høffding and Satne, 2019). In some cases, there may also be resonance between the musical group and the audience. These different situations do not entail autonomous high-church cognitive processes – as if what is required is a thinking or reflective contemplation. The performer does not think: "I'm in the concert hall playing with my quartet; therefore I should play in this style." It is rather that the concert hall and the people I am making music with elicit a specific feeling and style.

In some cases, a strong functional integration can be found in a form of musical joint attention, a shared sense of the music, a kind of entrainment and sensorimotor synchronization with the other players that produces a joint musical experience that approaches Merleau-Ponty's notion of intercorporeity.

> The intercorporeal inclusion of the other musician can be said to alter and expand the sense of agency, such that I no longer primarily attend egoically to my agency, my movements, my interpretation but see the entire setting, music, body, instrument, and even fellow musicians as one large agent. This is an affective and bodily we-intentionality: a musical intercorporeity or musical interkinesthetic affectivity (Høffding, 2019, p. 244).

The meshing of the horizontal and vertical axes may also involve "joint body schemas" in practices that have been shown to extend an individual's peripersonal space to include the other person, evidenced in changes to neuronal and behavioral processes (Soliman and Glenberg, 2014). As Soliman and Glenberg showed, these body-schematic effects are not simply modulated top-down by cultural practices, but rather, such social and cultural factors are incorporated into body-schematic processes which, in turn, express them in motoric performance. The situation of performance thus involves distributed and temporally extended processes that include all relevant variables – embodied, ecological, intersubjective/social, and cultural. These are not the accomplishments of narrow processes taking place just in-the-head, or strictly on a vertical axis, but are processes that extend into the world, meshed with the structures of our intercorporeal and material engagements.

Accordingly, there are multiple complex factors that extend on the horizontal axis and that shape the situation in which the agent is embedded. The notion of task dependency, where action may be defined by a particular role in the performance, is clearly relevant at this point, although it varies across different types of performance. Returning to the example of the standard tonal jazz performance mentioned in the first section, we find there clear normative task-dependent constraints that specify performance. Such constraints will vary across some proportion of functional integration and task dependency defining both vertical and horizontal meshing. Accordingly, how the "head," the solos, and the outro actually play out will in varying degrees depend on not just the vertical mesh of cognitive and intrinsic body-schematic control for each individual player but also on our interactions with co-players, on the music itself, and on affective modulations that may permeate the entire situation.

CONCLUSION: PERFORMANCE AND SITUATED COGNITION AS MUTUALLY ENLIGHTENING

We have argued that the notion of a meshed architecture can generalize beyond performance studies and contribute more broadly to an understanding of situated cognition. Once we understand that performance is not mindless, the model of a meshed architecture allows us to specify not only how cognition plays a role in performance but also how other factors situate performance. In regard to this model, we proposed four clarifications: (1) that the cognitive processes involved in performance are complex and varied and can include a full register that goes from explicit conscious control to implicit pre-reflective consciousness; (2) that control is not just top-down but can also be intrinsic to bottom-up processes; (3) that the mesh is complicated by the horizontal integration of ecological, social, and cultural/normative factors that extend beyond the performing agent but nonetheless constrain or contribute to performance; and (4) that affective factors are central for both modulating intrinsic control features and integrating the vertical and horizontal axes of the mesh. By making these clarifications, we have provided a more adequate view of how functional integration and task dependency work in situated cognition.

Accordingly, we think that the meshed architecture model of performance provides a way to explicate various processes involved in situated cognition and helps to make the concepts of functional integration (the coupling of agent and world) and task dependency (most typically defined by social, cultural, and

normative factors on the horizontal axis) less abstract. At the same time, notions of situated cognition that involve functional integration and task dependency help to enrich the conception of a meshed architecture.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

SG's research was supported by an Australian Research Council grant, Minds in Skilled Performance, DP170102987.

REFERENCES

Bernstein, N. A. (1984). "Some emergent problems of the regulation of motor acts," in *Human Motor Actions: Bernstein Reassessed*, ed. H. T. A. Whiting (Amsterdam: North-Holland), 354–355.

Berthoz, A. (2000). *The Brain's Sense of Movement*. Cambridge, MA: Harvard University Press. Trans. G. Weiss.

Christensen, W., and Sutton, J. (2019). "Mesh: cognition, body, and environment in skilled action – a new introduction to "cognition in skilled action."," in *Handbook of Embodied Cognition and Sport Psychology*, ed. M. L. Cappucio (Cambridge, MA: MIT Press), 157–164.

Christensen, W., Sutton, J., and McIlwain, D. J. (2016). Cognition in skilled action: meshed control and the varieties of skill experience. *Mind Lang.* 31, 37–66. doi: 10.1111/mila.12094

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

Clark, A., and Wilson, R. (2009). "How to situate cognition: letting nature take its course," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (Cambridge: Cambridge University Press), 55–77. doi: 10.1017/cbo9780511816826.004

Clarke, E., DeNora, T., and Vuoskoski, J. (2015). Music, empathy and cultural understanding. *Phys. Life Rev.* 15, 61–88. doi: 10.1016/j.plrev.2015.09.001

Cohen, R. (2013). *Acting Power*. London: Routledge.

Colombetti, G. (2014). *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge: MIT Press.

Dewey, J. (1922). *Human Nature and Conduct: An Introduction to Social Psychology*. New York: Modern Library.

Dreyfus, H. (2005). Overcoming the myth of the mental: how philosophers can profit from the phenomenology of everyday expertise. *Proc. Addr. Am. Philos. Assoc.*79, 47–65.

Dreyfus, H. (2007). Response to McDowell. *Inquiry* 50, 371–377. doi: 10.1080/00201740701489401

Fitts, P. M., and Posner, M. I. (1967). *Human Performance*. Belmont, CA: Wadsworth

Gallagher, S. (2013). The socially extended mind. *Cogn. Syst. Res.* 2, 4–12.

Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.

Gallagher, S., and Aguda, B. (2020). Anchoring know-how: action, affordance and anticipation. *J. Conscious. Stud.* 27, 3–37.

Gallagher, S., Mastrogiorgio, A., and Petracca, E. (2019). Economic reasoning in socially extended market institutions. *Front. Psychol.* 10:1856. doi: 10.3389/fpsyg.2019.01856

Gode, D., and Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders. *J. Polit. Econ.* 101, 119–137. doi: 10.1086/261868

Høffding, S. (2019). *A Phenomenology of Musical Absorption*. Cham: Palgrave MacMillan.

Høffding, S., and Satne, G. (2019). Interactive expertise in solo and joint musical performance. *Synthese* 1–19. doi: 10.1007/s11229-019-02339-x

Hutchins, E. (2014). The cultural ecosystem of human cognition. *Philos. Psychol.* 27, 34–49. doi: 10.1080/09515089.2013.830548

Jonides, J., Naveh-Benjamin, M., and Palmer, J. (1985). Assessing automaticity. *Acta Psychol.* 60, 157–171. doi: 10.1016/0001-6918(85)90053-8

Kirsh, D., and Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cogn. Sci.* 18, 513–549. doi: 10.1207/s15516709cog1804_1

Krueger, J. (2014). Varieties of extended emotions. *Phenomenol. Cogn. Sci.* 13, 533–555. doi: 10.1007/s11097-014-9363-1

Krueger, J. W. (2019). "Music as affective scaffolding," in *Music and Consciousness II*, eds R. Herbert, D. Clarke, and E. Clarke (Oxford: Oxford University Press), 55–70. doi: 10.1093/oso/9780198804352.003.0004

Legrand, D. (2007). Pre-reflective self-consciousness: on being bodily in the world. *Janus Head* 9, 493–519.

Legrand, D., and Ravn, S. (2009). Perceiving subjectivity in bodily movement: the case of dancers. *Phenomenol. Cogn. Sci.* 8, 389–408. doi: 10.1007/s11097-009-9135-5

Logan, G. D. (1985). Skill and automaticity: relations, implications, and future directions. *Can. J. Psychol. Rev. Canad. Psychol.* 39, 367–386. doi: 10.1037/h0080066

Maiese, M. (2018). Embodiment, sociality and the life-shaping thesis. *Phenomenol. Cogn. Sci.* 18, 353–374. doi: 10.1007/s11097-018-9565-z

Merleau-Ponty, M. (2012). *Phenomenology of Perception*. London: Routledge. trans. D. A. Landes.

Montero, B. (2010). Does bodily awareness interfere with highly skilled movement? *Inquiry* 53, 105–122. doi: 10.1080/00201741003612138

Montero, B. G. (2015). Thinking in the zone: the expert mind in action. *Southern J. Philos.* 53(Suppl. 1), 126–140. doi: 10.1111/sjp.12119

Montero, B. G. (2016). *Thought in Action: Expertise and the Conscious Mind*. New York: Oxford University Press.

Petracca, E., and Gallagher, S. (2020). Economic cognitive institutions. *J. Inst. Econ.* 1–19. doi: 10.1017/S1744137420000144

Ratcliffe, M. (2012). "The phenomenology of existential feeling," in *Feelings of Being Alive*, eds J. Fingerhut and S. Marienberg (Berlin: De Gruyter), 23–54.

Robbins, P., and Aydede, M. (2009). "A short primer on situated cognition," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (Cambridge: Cambridge University Press), 3–10. doi: 10.1017/cbo9780511816826.001

Rowlands, M. J. (2010). *The New Science of the Mind: From Extended mind to Embodied Phenomenology*. Cambridge, MA: MIT Press.

Ryan, K., and Gallagher, S. (2020). Between ecological psychology and enactivism: is there resonance? *Front. Psychol.* 11:1147. doi: 10.3389/fpsyg.2020.01147

Sadikot, S. (2014). *Mindlessness is Crucial in Batting: Sangakkara. BCCI: Board of Controlfor Cricket in India, 9 November. Hyderabad*. Available online at: http://www.bcci.tv/news/2014/features-and-interviews/8893/mindlessness-is-crucial-in-batting-sangakkara (accessed January 15, 2015).

Salice, A., Høffding, S., and Gallagher, S. (2017). Putting plural self-awareness into practice: the phenomenology of expert musicianship. *Topoi* 38, 197–209. doi: 10.1007/s11245-017-9451-2

Shusterman, R. (2008). *Body Consciousness: A Philosophy of Mindfulness and Somaesthetics*. Cambridge: Cambridge University Press.

Slaby, J., and Gallagher, S. (2015). Critical neuroscience and the socially extended mind. *Theory Cult. Soc.* 32, 33–59. doi: 10.1177/0263276414551996

Slors, M. (2019). Symbiotic cognition as an alternative for socially extended cognition. *Philos. Psychol.* 32, 1179–1203. doi: 10.1080/09515089.2019.1679591

Slors, M. (2020). From notebooks to institutions: the case for symbiotic cognition. *Front. Psychol.* 11:674. doi: 10.3389/fpsyg.2020.00674

Soliman, T. M., and Glenberg, A. M. (2014). "The embodiment of culture," in *The Routledge Handbook of Embodied Cognition* , ed. L. Shapiro (London: Routledge), 207–220.

Stanley, J. (2011). *Know How*. Oxford: Oxford University Press. Stormark & Braarud.

Sutton, J., McIlwain, D., Christensen, W., and Geeves, A. (2011). Applying intelligence to the reflexes: embodied skills and habits between Dreyfus and Descartes. *J,. Br. Soc. Phenomenol.* 42, 78–103. doi: 10.1080/00071773.2011.11006732

Tribble, E. B. (2016). "Distributed cognition, mindful bodies and the arts of acting," in *Theatre, Performance and Cognition: Languages, Bodies and Ecologies*, eds N. Shaughnessy and J. Lutterbie (London: Bloomsbury Publishing), 132–140.

Varga, S. (2019). *Scaffolded Minds*. Cambridge, MA: MIT Press.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 21 August 2020
doi: 10.3389/fpsyg.2020.01966

Attuning to the World: The Diachronic Constitution of the Extended Conscious Mind

Michael D. Kirchhoff[1]* *and Julian Kiverstein*[2]

[1] Department of Philosophy, University of Wollongong, Wollongong, NSW, Australia, [2] Department of Psychiatry, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands

It is a near consensus among materialist philosophers of mind that consciousness must somehow be constituted by internal neural processes, even if we remain unsure quite how this works. Even friends of the extended mind theory have argued that when it comes to the material substrate of conscious experience, the boundary of skin and skull is likely to prove somehow to be privileged. Such arguments have, however, typically conceived of the constitution of consciousness in synchronic terms, making a firm separation between proximate mechanisms and their ultimate causes. We argue that the processes involved in the constitution of some conscious experiences are diachronic, not synchronic. We focus on what we call phenomenal attunement in this paper—the feeling of being at home in a familiar, culturally constructed environment. Such a feeling is missing in cases of culture shock. Phenomenal attunement is a structure of our conscious experience of the world that is ubiquitous and taken for granted. We will argue that it is constituted by cycles of embodied and world-involving engagement whose dynamics are constrained by cultural practices. Thus, it follows that an essential structure of the conscious mind, the absence of which profoundly transforms conscious experience, is extended.

Keywords: extended consciousness, extended mind, cultural practices, diachronic constitution, ultimate explanation, proximate explanation

OPEN ACCESS

Edited by:
Beate Krickel,
Technical University of Berlin, Germany

Reviewed by:
Wanja Wiese,
Johannes Gutenberg University Mainz, Germany
Georg Northoff,
University of Ottawa, Canada

***Correspondence:**
Michael D. Kirchhoff
kirchhof@uow.edu.au

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *29 February 2020*
Accepted: *15 July 2020*
Published: *21 August 2020*

Citation:
Kirchhoff MD and Kiverstein J (2020) Attuning to the World: The Diachronic Constitution of the Extended Conscious Mind. Front. Psychol. 11:1966. doi: 10.3389/fpsyg.2020.01966

INTRODUCTION

In this paper, we set out to defend the thesis of the *extended conscious mind* (ECM). We do so because we take it that the mind in general is first and foremost widely and diachronically constituted. The vast majority of what humans think and experience unfolds over time through bouts of situated engagement with the environment. This does not just hold for unconscious problem solving as many leading exponents of extended mind theory are disposed to argue. These philosophers will happily concede that some of our unconscious cognition is accomplished by cycles of perception and action in which the cognizer makes active use of resources located in the environment around them (see, e.g., Clark and Chalmers, 1998; Clark, 2008; Menary, 2010; Sutton et al., 2010; Wheeler, 2010; Kirchhoff, 2012; Kiverstein, 2018). Most of these philosophers have, however, been unwilling to generalize such arguments to consciousness (see, e.g., Chalmers, 2008, 2019; Clark, 2009, 2012). They have argued that when it comes to consciousness the boundary of the skin and skull will somehow turn out to be privileged and special. Others have conceded that in some sense ECM is possible (Wheeler, 2015). But they have claimed that specific arguments for ECM have thus

far failed to make a convincing case that consciousness actually does extend. They claim that our best sciences of consciousness make it highly likely that consciousness will turn out to be a purely "in-the-head," brain-based phenomenon (Clark, 2009; Wheeler, 2015).

We argue, by contrast, that there are no good grounds for setting up a divide between unconscious cognition and conscious perceptual experience. What is good for the goose (extended unconscious cognition) is also good for the gander (extended conscious experience). The boundary of skin and skull has no special properties such that only the processes that fall within this boundary have what it takes to support conscious experience. The cognitive agent is what Susan Hurley called a *dynamical singularity*—one that forms out of a field of causal flows, some of which loop out into the world through cycles of perception and action (Hurley, 1998). Thus, the boundary of the conscious mind can, in the right kind of circumstances, form in an agent's dynamical coupling with its environment.

In what follows, we restrict our argument to a phenomenological structure of everyday lived experience we term "phenomenal attunement"—the feeling of being at home in a familiar culturally constructed environment. This phenomenological structure forms in the co-constituting coupling of the human agent with its social and cultural environment. We talk of the "co-constitution" of agent and environment because we will argue both agent and environment form together. The individual's cognitive capacities are partially constituted by environmental structures, practices, and institutions. At the same time, these structures, practices and institutions are the product of human cultural activities. There is no end point in the process leading to the experience of phenomenal attunement after which the individual can throw away the cultural environment and rely solely upon the brain. Since phenomenal attunement is a structure of conscious experience, this will provide us with an argument for why the person's conscious experience cannot always be generated solely out of processes unfolding inside the person's brain, uncoupled from the surrounding environment. The cultural environment plays a constituting role because we get to experience only phenomenal attunement (and its corollary of phenomenal disattunement) in our ongoing co-constituting coupling with a social and culturally constructed niche.

Internalist critics of ECM will be quick to insist (mistakenly, we believe) that internalism is entirely consistent with this line of argument. They will most likely object that the brain is causally dependent on specific forms of agent-environment couplings to settle on the pattern of neural activity constitutive of a particular conscious experience (see, e.g., Adams and Aizawa, 2001; Rupert, 2009). Internalists will concede that the world plays an ongoing *causal* role in driving the brain into a certain neural configuration allowing for the emergence of conscious experience. But they will insist it is the particular neural configuration in question that materially constitutes the conscious experience (see also Clark, 2009; Wheeler, 2015). They will thus take issue with our talk of co-constituting coupling of agent and environment.

We address this objection directly by arguing that it rests on a misunderstanding of the distinction between causation and constitution, treating one as strictly diachronic (causation) and the other as wholly synchronic (constitution). The constitution relation is generally cast as a strictly non-causal (i.e., atemporal/synchronic) one of dependence. We argue (building on Kirchhoff, 2015b and Kirchhoff and Kiverstein, 2019a) such an understanding of constitution while appropriate for material objects, is ill-suited when it comes to characterizing dynamic and processual phenomena such as conscious experience. To adequately characterize the constitution of a process, it is both possible and fruitful to understand the concept of constitution as a diachronic relation of dependence. The notion of diachronic constitution we argue leads naturally to an extended account of phenomenal attunement, incorporating both ultimate and proximate causes.[1]

The structure of the paper is as follows. In section "The Diachronic Constitution of Phenomenal Attunement: The Case of Culture Shock," we start by explaining what we mean by phenomenal attunement. We illustrate this phenomenon by reference to the cases of culture shock and psychopathology in which it is disturbed. We argue that culture shock shows how the experience of being attuned to the cultural environment is an integral part of the phenomenology of our everyday conscious experiences. But phenomenal attunement is also constitutively dependent on the ongoing coupling of an individual to her cultural environment through cycles of perception and action. Thus, phenomenal attunement provides us with a case that illustrates how coupling to the cultural environment diachronically constitutes a core dimension of conscious experience. Section "Assembling the Mind: Cognitive Assembly and The Pac-Man Intuition" takes up a likely internalist objection to our argument. Arguments for the extended mind have tended to limit bouts of extended cognition to short, synchronic timescales. We argue that this focus on the synchronic is problematic, as it precludes dynamical processes unfolding over longer periods of time, from being more than ultimate (background) causes against which the brain assembles the elements that make up extended minds. We propose a new metaphysics of constitution, cast in terms of diachronic constitution that avoids this consequence. In section "Synchronic and Diachronic Constitution," we review the standard notion of constitution (synchronic constitution), which we then contrast with the diachronic conception of constitution required for understanding the constitution of dynamic processes. We suggest that the diachronic conception of constitution is required

[1]Others have defended positions on consciousness closely related to the extended mind view we develop in this paper. Ward (2012), for example, does so by appealing to personal-level considerations about the phenomenology of experience and what such considerations might tell us about the sub-personal level machinery involved in the constitution of consciousness. Others, like Noë (2004, 2009), appeal to active sensorimotor engagement with the environment to make their case (see also Hurley, 1998; Rowlands, 2010). Cosmelli and Thompson (2010) have argued for the constitutive dependence of consciousness on the non-neural body and world by calling into question brain-in-vat intuitions. Northoff (2019) can also be read as arguing for an account of consciousness as partially constituted by factors beyond the brain. Our argument will, however, differ from these important exponents of ECM. We argue for the constitutive dependence of conscious experience on cycles of perception and action that couple a person to his or her cultural environment. We base our argument for the extendedness of phenomenal attunement on a diachronic account of constitution.

to account for the metaphysics of extended minds. This is because extended cognition is dynamic, unfolding over time through cycles of situated engagement with the affordances or possibilities for action the environment furnishes (Anderson et al., 2012; Kiverstein, 2018). In section "Objections: Pluggability Intuitions, Free-Floating Brains and Internal Fantasies," we provide responses to three objections against our argument for ECM. These objections aim to defend the consensus view among materialist philosophers of mind that all experiences must be somehow constituted out of internal neuronal processing. In section "Wide and Diachronic Constitution: Two Conceptual Flips," we tackle the often made objection that arguments for the extended mind (EM) are guilty of conflating causal coupling with the metaphysical relation of constitution. We argue that once one makes the turn to diachronic constitution, this objection against EM, and by extension ECM, loses its force. We end section "Wide and Diachronic Constitution: Two Conceptual Flips," by showing how the diachronic view of constitution we argue for in this paper can safely avoid the cognitive bloat objection often raised against EM (see, e.g., Rowlands, 2009; Sprevak, 2009).

THE DIACHRONIC CONSTITUTION OF PHENOMENAL ATTUNEMENT: THE CASE OF CULTURE SHOCK

As an illustration of what we mean by phenomenal attunement, we will begin by considering the example of culture shock. An experience of culture shock is characterized by feelings of distress and alienation. These feelings of distress and alienation are examples of an absence of phenomenal attunement with the cultural environment. A much-discussed case is 13-year-old Eva Hoffman, who, along with her mother and father, left Poland in 1956 for the prospects of a better life in Vancouver, Canada. Even though Eva had her parents by her side, her experiential world changed dramatically. She explains:

> *[T]he country of my childhood lives within me with a primacy that is a form of love . . .. It has fed me language, perceptions, sounds. . .. It has given me the colors and the furrows of reality, my first loves (Hoffman, 1989, pp. 74–5; quoted in Wexler, 2008, p. 175).*

Having spent only three nights in Vancouver, she reports waking up from a dream, wondering:

> *[W]hat has happened to me in this new world? I don't know. I don't see what I've seen, don't comprehend what's in front of me. I'm not filled with language anymore, and I have only a memory of fullness to anguish me with the knowledge that, in this dark and empty state, I don't really exist (Hoffman, 1989, p. 180; quoted in Wexler, 2008, p. 175).*

Culture shock illustrates how expectations that have their origin outside of the individual in patterns of cultural practices attune us to a shared cultural environment. Should the individual move to a new environment, the result may be misalignment and pervasive, hard-to-suppress violation of her expectations about her shared social and cultural environment. We will argue that to properly explain cases such as culture shock we need to appeal to an extended dynamic singularity comprising Eva's internal neurobiological states, the patterns of practice that are enacted within her cultural niche, and her sensory and active states that couple her to her cultural environment. To explain her current experiences one must take into account the expectations that she has formed through her past involvement in cultural practices and the role of these expectations in shaping the phenomenology of her ongoing experience. It is not just her past that we need to take into account but also her present circumstances and her orientation to the future in her new cultural environment.

Phenomenal attunement can be formalized as the divergence between prior expectations about the causes of sensory observations and the actual causes (e.g., generated via patterns of cultural practice). The experience of phenomenal attunement can be described as the Kullback–Leibler divergence between prior expectations (P^*) and cultural practices (Po) generating sensory states: $C_{\text{exp}} = D_{\text{KL}}\ [P^* \,||\, P\text{o}]$.[2] Phenomenal attunement comes in degrees and varies with divergence between P^* and Po. There is an experience of phenomenal lack of attunement when $D_{\text{KL}}\ [P^* \,||\, P\text{o}] > 0$. On average, one would expect expectations to converge on cultural practices, ensuring phenomenal attunement to one's cultural environment. Suppose, now, that we associate experiences of culture shock (feelings of distress and alienation) with uncertainty; then the higher the divergence between P^* and Po, the more uncertainty is expressed in the coupling between P^* and Po. Crucially, if there is high uncertainty as a consequence of the divergence between P^* and Po, the subject will need to exert more effort to make sense of her surroundings. If one's expectations systematically fail to align with the regularities (causal and statistical) of one's environment, feelings of distress are likely to arise as one needs to make much more effort to make sense of how one finds oneself situated in the world.

We are all familiar with such situations, where uncertainty about outcomes of social interactions yield sensations of frustration or discontent. Alignment and continued attunement to other people and to wider patterns of practice are integral parts of the phenomenology of our everyday conscious experiences. Slaby (2016) invites us to imagine working as an intern in a large company:

> *Your first days working in the firm will be marked by experiences like the following: You find the regular employees speaking, acting, moving, and comporting themselves in ways that are unfamiliar to you in various ways. Not only will their work routines be new to you, but also their styles of interacting, of comporting themselves, of resonating affectively with one another, the ways of address, of conversing with superiors, the use of humor to begin a conversation, or deflate a moment of tension, when and how to*

[2]KL-divergence is a part of the formal tools used for modeling consciousness in terms of belief or expectation updating schemes such as predictive processing in cognitive neuroscience (Hohwy, 2012; Clark, 2016; Friston, 2018; Kirchhoff and Kiverstein, 2019a). It might be objected that this alignment between P^* and sensory states with cultural practices can be explained from entirely inside of the brain. The remainder of our paper is devoted to explaining why we think such an objection is mistaken. See also Northoff (2019) for an account of why KL-divergence (which he understands in terms of variational free energy) is best interpreted in environment-involving terms (and Kirchhoff and Kiverstein, 2019a for a book-length treatment of the extended conscious mind cast in terms of active inference and predictive processing).

display certain feelings openly (enthusiasm maybe, or pride after an achievement), or suppressing others (no fear, no insecurities), and so on (Slaby, 2016, p. 1).

As an intern you initially experience a phenomenal lack of attunement to others in the workplace. So much of what takes place between colleagues in a large corporation is the enactment of a past history of interaction to which outsiders are not privy. To align and adjust to this novel niche, an intern will need to become familiar with what other employees take for granted. She must learn more than how to perform her work routine. She must cultivate a sense of how to interact with her colleagues and what is at issue in these interactions. As long as she does not have a sense of this, she will experience just the same or similar feelings of distress and alienation associated with culture shock. Perception and action in social domains such as a large corporation are organized by norm-regulated practices—regular, stable, and ordered patterns of activity. Norms structure social interactions within a social domain. One must become attuned to the norms that govern interactions in domains of social life in order to feel at home in these domains of social life.

Thus, maintaining attunement with one's cultural surroundings critically depends on a match in an individual's expectations and the normatively regulated expectations of other participants in a social practice (Kirchhoff and Kiverstein, 2019a, ch. 5).[3] The expectations that guide one's perception and action must match the expectations that form in patterns of practice. Pervasive and sustained lack of attunement can prove to be pathological. People suffering from schizophrenic delusion have been hypothesized to have a high expectation of noise and uncertainty. They expect more sensory noise than there really is in a given context with the consequence that they are unable to find the signal among the noise. This leads them to neglect sensory evidence in favor of their prior expectations (Fletcher and Frith, 2009; Hohwy, 2015). The effect of this aberrant weighting of evidence and general failure of context-sensitive updating of prior expectations is that they come to inhabit a delusional reality that is increasingly cut off and removed from the common-sense, everyday, familiar reality they share with other people (Sass, 1994). They increasingly come to inhabit their own solipsistic reality. People with autism by contrast give too much weight to new sensory evidence. This weighting of sensory evidence leads to sensory information that conflicts with their prior expectations to dominate in processing, which has the consequence that they have difficulties in becoming attuned to more stable and persistent regularities (Pellicano and Burr, 2012; Lawson et al., 2014; Palmer et al., 2017). The consequence of this aberrant weighting of sensory information in both cases is that people have difficulties in becoming phenomenally attuned to non-autistic cultural practices.[4]

In culture shock this attunement to the everyday world is also temporarily lost, and this leads to a deep disturbance of lived experience with the divergence between P^* and $P\mathrm{o}$ being high. Crucially, attunement, as well as lack of attunement, relies on the ongoing coupling of Eva to the cultural world through cycles of perception and action. Phenomenal attunement is, as we have suggested, the outcome of synchronous coupling between internal and external states mediated via sensory and active states (c.f. Kirchhoff and Kiverstein, 2019a). The person is coupled to her cultural environment by sensory and active states. Patterns of practice structure what we expect to experience in our cultural environment. Repeated engagement in these practices establishes the norms and rules of conduct that push back should one deviate from them. Think of Slaby's example of the workplace practices and patterns of micro-interaction in a large corporation. Roepstorff et al. (2012) state: "Culture gets under the skin and skull, . . . and it is remade gradually through collective instances of actualization" (p. 1052). This normatively regulated coupling plays a constituting role in the generation of conscious experience of phenomenal attunement or lack of attunement.

Phenomenal attunement is not constituted synchronically by underlying brain states at a snapshot instant in time. If conscious experience is equal to the Kullback–Leibler divergence between prior expectations (P^*) and cultural practices ($P\mathrm{o}$) generating sensory states, $C_{\mathrm{exp}} = D_{\mathrm{KL}}\,[P^* \,||\, P\mathrm{o}]$, then attunement cannot be constituted by the proximate mechanisms involved in generating P^*. The KL-divergence is a *relational measure* of the distance between P^* and $P\mathrm{o}$. An individual agent's coupling to a culturally constructed environment is best understood diachronically, not synchronically. The individual's activities are constrained by the norms that govern the regular and ordered patterns of activities that stabilize in a community over time. Thus, the experience of phenomenal lack of attunement that arises from failing to become adequately attuned to the cultural environment is best understood in terms of dynamical processes that unfold over multiple interacting timescales. In other words, it is best understood diachronically. Let us unpack this argument step by step.

We have analyzed Eva's experience of culture shock in terms of expectations P^* formed out of her involvement in past practices associated with her childhood in Warsaw that fail to align with the cultural environment she now inhabits in Vancouver. The cycles of perception and action that couple her to her cultural surroundings are "permeated" and "infused" by her expectations (Gallagher, 2018; Hutto et al., 2019). Her expectations provide her with a background understanding of her surroundings in virtue of which she encounters a familiar environment in which she knows how to act. Think again of our example of the office

[3] In neural dynamical terms, we can think of attunement with respect to generalized synchrony of the large-scale internal dynamics of the brain with the external dynamics of the cultural environment. A simple example of generalized synchrony is entrainment of the sort that happens when one finds one tapping one's foot along to the rhythm of music one is listening to. Northoff (2019) suggests that in perception the temporal-spatial dynamics of the brain synchronize with, and thus conform to, the world's external dynamics, while in action it is the other way around: the world's external dynamics conforms to the brain's internal dynamics (c.f. Bruineberg and Rietveld, 2014; Bruineberg et al., 2018; Kirchhoff and Kiverstein, 2019a, ch. 4). Northoff makes a more general argument for ECM on the basis of these kinds of considerations. We restrict our focus here to arguing for the extension of phenomenal attunement.

[4] Krueger and Maiese (2018, p. 28) have noted that a key challenge people on the autism spectrum face is becoming attuned to the norms and expectations of non-autistic "neurotypicals," which are often "unspoken, highly context-specific and communicated by way of nuanced body language." They go on to write that "high-functioning" people with autism often find it easy to become attuned to each other since their interactions are governed by "autism-friendly" norms and expectations.

intern who is yet to be initiated into the styles of interacting taken for granted by other employees. Eva's expectations, however, fail to align with the normative expectations regulative of people's activity in Vancouver. It is these expectations that she brings to bear to make sense of her present situation and to orient how she engages with her surroundings in the future. However, they fail to orient her adequately to the normative expectations operative in her current environment. The result of this lack of alignment is her experience of phenomenal lack of attunement.

Those in the grip of internalist intuitions might agree with us that patterns of cultural practices are involved in setting the parameters of brain-based processes over long (ultimate) timescales. Yet they will insist that in the here and now, conscious experience is determined entirely by proximate mechanisms in the brain unconsciously inferring hypotheses about the hidden external causes of sensory data. To insist otherwise would be to fallaciously confuse ultimate causes with proximate ones in an explanation of consciousness. Our objector might attempt to bolster such intuitions by invoking neural duplicates. Would Eva's neural duplicate in the present moment not have the same phenomenal experience as Eva just now? This objection turns on the idea that it is the neural machinery and the particular forms of information processing it supports that do the work of constituting the conscious mind synchronically, not a history of interaction with the environment. We show why we think such an objection is misplaced in the next section.

ASSEMBLING THE MIND: COGNITIVE ASSEMBLY AND THE PAC-MAN INTUITION

The debate about the extended mind (EM) really took off in the philosophy of mind with the publication of a short paper in *Analysis* by Clark and Chalmers (1998). The aim of the paper was to invite readers to question the (biological chauvinist) assumption that anything located external to skin and skull cannot be a part of a person's mind. Clark and Chalmers (1998) devised much discussed thought experiments aimed at showing how something is a part of a person's mind because of the causal role it plays in guiding the person's behavior. Artifacts, such as notebooks located outside of the individual's body, can become fully integrated parts of an individual's thinking processes. By coupling with tools and technologies the individual can accomplish her thinking and problem solving. Thus, artifacts can form a part of a larger cognitive system the individual relies upon in acting. They can come to play a constitutive role in the production of the individual's behavior equivalent to that played by processes internal to the individual. Clark and Chalmers (1998) argued that we should not exclude the things around us from counting as parts of our minds simply because of their location outside of the head; rather, if external elements play the right kind of functional role in driving cognitive processes, such elements should count as part of someone's mind—just as internal states playing such roles would naturally qualify as part of one's mental machinery.

Clark in his later work was up-front about giving a privileged place to the agent in the assembly or formation of extended cognitive processes (Clark, 2008). He writes: "Human cognitive processing (sometimes) literally extends into the environment surrounding the organism. But the organism (and within the organism, the brain/CNS) remains the core and currently the most active element" (p. 139). The individual cognizer decides, in part based on efficiency considerations, whether to rely solely on her own on-board (neural) cognitive machinery to solve a problem or to softly assemble a solution that makes use of resources located in the external environment. The work of assembling a cognitive system that can solve a particular problem is delegated to the brain of the individual. Problem solving may sometimes constitutively involve bouts of situated, real-world action that unfolds over relatively short timescales of hundreds of milliseconds, or seconds, and is orchestrated from inside of the brain of the individual. Insofar as Clark takes cognition to be organism-centered, he must insist upon a strict separation of events as they unfold over short timescales from events as they unfold in cultural practice over longer historical timescales. But many examples of situated action in the literature are examples of actions the person has learned to perform by taking part in cultural practices (Hutchins, 2011). Whereas Clark will argue it is the brain that does the work of assembling and organizing the cognitive system in these cases, we would argue (taking our lead from the cognitive anthropologist Ed Hutchins) that in many cases cultural practices organize the action in situated action and therefore in cognitive assembly.

The patterns of perception and action belong to a cultural practice because the understanding of what to say and what to do derives from rules, evaluative standards, principles, and imperatives that are operative in practice. Practices organize what people do in the sense that the tasks and projects people undertake and the purposes and ends for which people act have their origin in practices (Schatzki, 1996). What the members of a practice say and do follows from and aligns with the practice. Think again of Slaby's example of the intern working in a large corporation. Employees are trained to think, feel, and act so that they can become attuned to playing a specific role within the corporate machine. People already habituated to working in the company will reinforce and sanction or punish what the intern says and does in more or less subtle ways until what she says and does is well aligned with the prevalent styles of interaction in this institution. Individuals are situated in practices, but the practices also situate what individuals do and say. Clark and Chalmers (1998) put forward the hypothesis of the extended mind starting from a picture in which a pre-existing individual agent occasionally connects with the world to solve a problem that it would be much harder to solve without the use of some artifact, tool, or technology. We are arguing for a view of the extended mind in which the activities of the individual agent and the agent's cultural environment are quite literally co-constituting. The individual isn't already fully formed but what the individual says and does is profoundly shaped and transformed by the practices they take part in.

Clark is willing to allow that cultural practices may do some of the work of *setting the scene* for the assembly and orchestration of extended cognitive processes. However, he stops short of allowing cultural practices to form a part of the slow unfolding processes

out of which extended minds form. Clark concedes what no doubt everyone will allow—that a child must have learned to read and write before she can make use of pen and paper to do multiplication. However, he argues that when the child makes use of the external scaffolding of pen and paper to do long multiplication, she does so over relatively short time scales. But why does Clark privilege processes unfolding in the synchronic here and now? What the child is doing in making use of pen and paper is reenacting what she has learned by taking part in a practice. The actions she performs are embedded in and organized by the practice of which she is a part. History and culture are always embedded as well as carried along in the practices and artifacts individuals are engaging with (Menary, 2007, 2010; Sutton, 2010; see also Haugeland, 2002). The result of focusing only on the synchronic timescale—i.e., on proximate causes—is that everything that makes a difference outside of the here and now must be treated at best as making a causal contribution to mentality, either as background conditions or as input to internal neural processes.

Clark's reasoning has the problematic consequence that minds must be fully constituted over short-term timescales.[5] History and culture form background conditions that set the stage for the brain to do the real work of constituting the mind in the here and now. This wrongly assumes that all of the work of cultural practices in constraining, coordinating, and self-organizing action can come to be fully internalized. Clark seems to assume that what is learned from others through training in social practices can simply be internalized in the form of internal representations. This training can then get to do its work through its internal representation by the individual. The cultural transmission of knowledge and practices is understood as transmission of information among individuals. Once the information has circulated in the right way among individuals, there is no longer any work left for cultural practices to do.

We think this is the wrong model of how cultural learning works. To see what is mistaken in this picture, consider by way of analogy an individual we will call Pac-Man, named after the character in the arcade game. Evolution has set up Pac-Man so that on average and over time he distills the regularities of his niche. He "eats" up such regularities and comes to embody them in an internal model of his external environment. Pac-Man moves about his environment, extracting and consuming statistical structures to build up a detailed internal model of his local environment. This is how he learns about his niche. The body of Pac-Man and the wider niche in which he is situated are ultimately important for acquiring and updating the parameters of his internal model. Yet once these parameters are acquired, Pac-Man can rely on his internally encoded model of his world to act adaptively in his environment. He has consumed all of the information he needs. We will call this the Pac-Man intuition.

The Pac-Man intuition is false. We suggest by contrast that extended minds are constituted by temporally unfolding processes, and thus the Pac-Man intuition provides the wrong model for thinking about the internalization of cultural forms of knowledge. Internal models as they are embodied in living beings are tasked with always having to maintain a grip on the fluctuations in the dynamics of their local environments. The fluctuations do not reside or disappear but are constantly forming and reforming, even if in only slightly different ways. Organisms must therefore constantly *attune* their internal dynamics to the continuously changing dynamics of the environment in which they are situated. But now it might be objected this attunement takes place in the here and now. Past learning sets up dispositions to act in ways that conform with a practice. Think again of the child learning to do long multiplication. These dispositions are fully internalized. Everything that is required for the disposition to be realized in action happens synchronically in interaction with the environment and with other people, such as teachers. But to say that a disposition is internalized is not at all the same as saying that what people know when they take part in cultural practices is fully internalized. Thus, one can think of the enactment of cultural practices as happening synchronically without relying on the Pac-Man intuition to account for cultural learning.[6]

In reply, we suggest that this synchronic account of the enactment of cultural practices misses an important feature of situated action. It misses how the person's dispositions are constrained by rules, norms, principles, and standards that operate at the scale of the cultural practice. Situated action is constituted by processes that unfold over two timescales. First, there is the timescale of the cycles of perception and action that couple the agent to the environment as in the classic example of using pen and paper to do mental arithmetic. Cycles of perception and action unfold over time and thus cannot be synchronically constituted at a time *t*. As dynamical processes, they are diachronically constituted. Second, there is the slower timescale of the cultural practices the child is initiated into in learning to do long multiplication. The dispositions the child puts into action in the here and now are constrained by what people have done over the longer period of time during which the practice of doing multiplication has taken shape and developed. It is these timescales that get out of sync with each other in cases of phenomenal attunement as was argued at the end of the previous section. The expectations that are formative for Eva due to her growing up in Poland do not align with those that are operative in her new home of Vancouver. Thus, the expectations that shape her perception and action are out of step with those of her surroundings.

One can think of this entanglement of the slower and faster timescales by comparison with the dynamics of self-organizing systems—disordered systems in which global order can arise under the influence of the system's own dynamics. We observe the emergence of global order in such systems when a control parameter reaches a critical value that makes possible new forms of organization. Consider, for instance, the example of the

[5]Note, we are not questioning that problem solving may take this form, assembled and unfolding over synchronic timescales in the here and now. The point we are objecting to is that by limiting the assembly of even extended minds to synchronic timescales, one thereby rules out, unnecessarily, cultural practices unfolding over longer timescales from playing a role in the material constitution of what people say and do. A core aspect of our argument for ECM is precisely to question this privileging of the synchronic timescale.

[6]Our thanks to the reviewer for pressing this objection.

Bénard effect from non-equilibrium fluid dynamics. A Bénard or convection roll forms when a fluid (for example, oil) is heated from below. The temperature difference between the surface and the bottom of the fluid is the control parameter. Once this temperature gradient reaches a critical value with more energy being introduced into the fluid than can be dissipated, the fluid becomes unstable. This instability leads to the formation of rolling, convection patterns in the oil. These rolling patterns are macroscopic states of the fluid that slowly form in the oil as it is heated. Such a macroscopic state is formed by the molecules of which the oil is composed. Thus, there is a constraint that runs from the micro- to the macro-scale. But crucially, the constraints also run in the other direction from the macro- to the micro-scale. When the order parameter reaches a critical value, the system enters an unstable state that allows for the convection rolls to arise. The dynamics evolving over longer timescales—the temperature gradient over the ensemble—entrains the dynamics evolving over shorter timescales—the molecules and the dissipation of energy by the fluid.

We are suggesting just the same circular causal dynamic obtains in the case of situated action. The cycles of perception and action that form over relatively short timescales can be compared to the microscopic interactions that take place in the fluid when it is heated. We are suggesting that the rules, principles, and standards—the patterns of cultural practice—can be thought of as macroscopic-order parameters that evolve over longer timescales. These patterns of practice as order parameters form out of the interactions of individuals over time. But crucially, they also entrain what individuals do over faster timescales. The cycles of perception and action that couple the individual to the cultural environment and the patterns of practice that are up and running in the cultural environment mutually constrain each other. They form a circular causal relationship.

To attempt to account for situated action synchronically, just in terms of what happens here and now, is mistaken on two grounds. It ignores how the coupling of the agent to the environment in perception and action is a dynamic process that unfolds over multiple interacting timescales. Second, it abstracts away from the wider pattern of practice that is a constraint on the situated actions people perform over shorter timescales. The cultural "training wheels" cannot, always and necessarily, be dispensed with as the Pac-Man intuition implies. This is to assume, as Hurley (2010) has pointed out, "that extended tuning and maintenance processes" are no part of the sought-for explanation of the workings of the mind (Hurley, 2010, p. 142). We've argued against such an assumption. Once the Pac-Man intuition is rejected, however, we will need a different account of constitution from the one that assumes the mind can be constituted at a synchronic instant in time. We need a diachronic concept of constitution.

SYNCHRONIC AND DIACHRONIC CONSTITUTION

To introduce and develop the distinction between synchronic and diachronic constitution, a useful starting point is to get clear about the notion of a metaphysical grounding relation, of which the concept of constitution is one example. What characterizes a metaphysical grounding relation is the idea that for a relation, *R*, to qualify as a metaphysical grounding relation, *R* must express the form "*X* (or the *X*s) metaphysically determines *Y*," when it is *by virtue of* X (or the *X*s) that *Y* exists. Thus, in the context of our paper, *Y* is the experience of phenomenal attunement or its absence in cases of culture shock. We have been arguing that phenomenal attunement is constituted by cycles of perception and action that couple the perceiver to their local cultural environment. Thus, we are claiming that it is by virtue of the person's coupling to the cultural environment that the person has an experience of phenomenal attunement.

This *by-virtue-of* relation is often specified as a species of determination (cf. Kim, 1990; Shapiro, 2004; Polger, 2010). Different relations—such as composition, realization, and supervenience—have also been used in philosophy to express the view that *Y* exists by virtue of *X* (Kim, 1998; Bennett, 2011).

It is widely agreed that a necessary condition for *X* (or the *X*s) to constitute *Y* is that the relation of constitution that holds between X (or the Xs) and Y is a *synchronic* one-to-one, or many-to-one, relation of determination between spatially and materially co-located objects (or processes) of different kinds. A central reason for conceiving of constitution as a synchronic dependence relation is nicely articulated by Bennett: "building [grounding] relations do not unfold over time …. Causation, in contrast, is paradigmatically diachronic, and that idea is frequently invoked to distinguish causation from relations such as composition, constitution, supervenience …" (Bennett, 2011, pp. 93–94). The assumption that constitution must be a synchronic dependence relation is engrained in the very manner in which this grounding relation is analyzed. For example, it is a standard assumption on the part of constitution theorists that constitution requires spatial and material coincidence—X constitutes Y at t only if X and Y have the same spatial location at a particular time *t* and share the same material parts at that specific time *t*. It is thus presupposed that the constitution relation holds instantaneously between X (or the Xs) and Y and therefore cannot be a temporally unfolding relation. Causation, by contrast, may be said to hold between independent events or processes, in the sense that depending on the time interval between cause and effect, it is prima facie possible to think of cause as preceding its effect in time thus as occurring non-simultaneously. The standard formulation of constitution is thus that constitution is a synchronic relation of dependence.

It is not difficult to provide textual evidence for the claim that EM is typically taken to be a thesis about the constitution of minds that assumes this standard formulation of constitution (or, some other kind of metaphysical grounding relation):

> *EM is a claim about the composition or constitution of (some) mental processes (Rowlands, 2009, p. 54; italics added).*

> *What is at issue, as far as the claims about cognitive extension are concerned, is simply which bits of the world make true (by serving as the local mechanistic supervenience base for) certain claims about a subject's here-and-now mental states or cognitive processing (Clark, 2008, p. 118; italics added).*

> *Causal dependency of mentality on external factors—even when that causal dependency is of the "necessary" kind [. . .]—is simply not enough for genuine cognitive extension. What is needed is constitutive dependence of mentality on external factors, the sort of dependence indicated by talk of the beyond-the-skin factors themselves rightly being accorded fully paid-up cognitive status (Wheeler, 2010, p. 246; italics added).*

As a final example consider how Shapiro characterizes the difference between causation and constitution: "[If] *C* is a constituent of an event or process *P*, *C* exists where and when that event or process exists. Thus, for some process *P*, if *C* takes place prior to *P*'s occurrence [. . .], or if *C* takes place apart from *P*'s occurrence [. . .], then *C* is not a constituent of P" (Shapiro, 2011, p. 160).

The metaphysics of the extended mind has thus taken for granted that constitution is a synchronic relation of determination. But is this assumption warranted? Synchronic relations are not well suited for understanding dynamical processes or their nested or hierarchical organization. But candidates for cases of extended cognition typically involve reciprocal coupling of the agent and its environment. More formally, the equations describing the behavior of the agent over time cannot be solved independently of the equations describing the environment and vice versa (Lamb and Chemero, 2018). The variables in the respective equations describe how the components of the agent and environment change in relation to each other. The equations describing change in the environment contain variables whose values correspond to changes in the agent. Conversely the equations describing change in the agent contain variables whose values correspond to changes in the environment (Anderson et al., 2012). The state changes of the agent will be dampened and amplified by state changes in the environment and vice versa. The solution to these equations is thus interdependent.

Diachronic constitution captures the basic idea that for a process to be what it is, it must unfold over time. In other words, there is no such thing as a process at an instant or synchronic point in time. For example, one often reads that water is constituted by or composed of H_20. This assumption is a practical assumption to make, in science as in everyday life. But it should not be taken as evidence for the further claim that water is constituted by H_20 at a synchronic point in time. Instead water is constituted by "oxygen and hydrogen in various polymeric forms, such as $(H_2O)_2$, $(H_2O)_3$, and so on, that are *constantly forming*, *dissipating*, and *reforming* over short time periods in such a way as to give rise to the familiar properties of the macroscopic kind water" (Ladyman and Ross, 2007, p. 21; italics added). Hence, it makes "no sense to imagine it [water] having its familiar properties synchronically" (Ross and Ladyman, 2010, p. 160). Spivey (2007) makes the exact same point in his book-length treatment of cognitive processes and their underlying mechanisms.

We suggest that conceiving of constitution as a diachronic relation that unfolds over time makes a better fit with the extended mind in which a person's mental states form in the dynamic coupling of the agent with its surroundings. Diachronic constitution questions a basic assumption of synchronic constitution that ultimate causes must be treated as wholly distinct from proximate constituents. For example, we can ask why birds migrate, and we can ask how they migrate. The former question can be answered by reference to the evolutionary and developmental history of birds. The latter *how*-question is answered by reference to muscle mass, morphology, and so on, *at a specific point in time*.[7] So there is a clear temporal difference between these two forms of explanation: the ultimate explanation is a diachronic mode of explanation; whereas the proximate explanation, it might be thought, is a type of synchronic explanation. In other words, proximate *how*-explanations are mechanistic and immediate, while ultimate *why*-explanations are causal and historical. Diachronic constitution implies that the choice between ultimate and proximate explanations is sometimes a false choice. In the case of dynamical phenomena, proximate explanations will often include ultimate causes. The constitutive basis of a dynamical process or event will include both proximate mechanisms and processes unfolding over longer timescales (**Figure 1**).

Our claim that diachronic constitution integrates ultimate and proximate causes seems to obscure the distinction between causation and constitution. Must there always be a distinct way of identifying causes and constituents? We think not. A common strategy by which to identify constituents for specific phenomena is by determining what plays the most salient *causal role(s)* with regards to the constitution of some phenomenon. So the relevant distinction is not between causation and constitution, *per se*; rather, it is between *mere causes* and *constituent causes* (**Figure 1B**). Diachronic constitution as an account of the constitution of dynamic phenomena will include ultimate causes among its constituent causes.[8]

We conclude then that conceiving of constitution in diachronic terms provides the best metaphysical tools for understanding the kind of temporally nested structure that dynamical systems and processes exhibit (c.f. Kirchhoff, 2015a,c). Extended cognitive processes are constituted by many different subprocesses, each of which unfolds continuously over time exhibiting its own rate of change, rhythm and duration. Each process will be both influencing, and influenced by, the other subprocesses of which it is composed (van Gelder and Port, 1995). The constituent subprocesses may partially overlap in time but in order to contribute to the constitution of a system *S* it is not necessary that their existence entirely overlaps with that of *S*. The subprocesses that make up the agent-environment system do so over a temporally extended interval, and not at discrete instants

[7]Note that even answering this *how*-question presupposes a diachronic notion of constitution, for the processes involved in enabling flight are themselves temporally extended processes. To say that a system is in some particular state *X* at a particular point in time is to say that the average of the system's states during that period of time was *X* (see Spivey, 2007, for discussion). Thus, we do not agree that *how*-explanations must always posit mechanisms whose workings are synchronic. We are suggesting that such an assumption does not hold in the case of dynamical phenomena.

[8]One might worry that our account inherits the problem of how to delineate between mere ultimate causes and ultimate causes as constituent causes. We return to this worry below in section "Wide and Diachronic Constitution: Two Conceptual Flips."

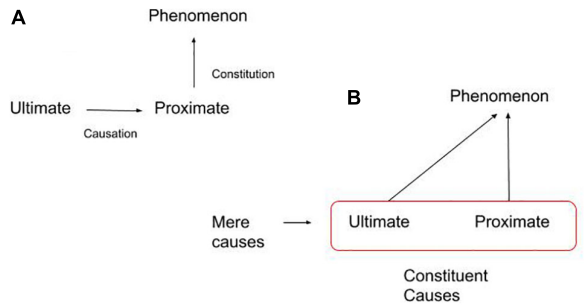

FIGURE 1 | In panel **(A)**, we depict the standard way of casting the difference between causation (ultimate causes) and constitution (proximate mechanisms), where ultimate causes influence but do not constitute the phenomenon of interest. Panel **(B)** illustrates the notion of diachronic constitution, where the constituent causes for some phenomenon include ultimate and proximate causes over time, but not mere (background) causes.

in a stepwise and linear manner. In the remainder of our paper we make use of the notion of diachronic constitution to develop and defend an extended account of phenomenal attunement.

OBJECTIONS: PLUGGABILITY INTUITIONS, FREE-FLOATING BRAINS, AND INTERNAL FANTASIES

Consider the following twin case involving Eva and Eva*. Eva and Eva* are neural duplicates. Eva, situated in Vancouver, is experiencing culture shock. Given that Eva* shares an identical neural profile with Eva, does it follow that she must also experience culture shock? We agree with Hurley (2010) that if a brain is not hooked up and plugged into an environment just like mine, it will not always be possible for this brain to play its role in generating experiences just like mine (c.f. Nöe, 2006). It is not always possible to "unplug" the internal neural factors that bring about experiences from the environment and "replug" them into a different environment without this replugging changing the functioning of the internal neural factors (Hurley, 2010). We'll argue that phenomenal attunement counts as a case in which pluggability fails. For Eva and Eva* to be neural duplicates they must also be environmental duplicates. It is not their being neural duplicates that minimally suffices to make them phenomenal duplicates.[9] The minimally sufficient conditions for Eva and Eva* to be phenomenal duplicates will include the cultural environment they couple to in perception and action and the normative expectations that operate this environment. This is because it is the coupling of Eva to the cultural environment and the role of culture-specific expectations in mediating this coupling that constitute her experience of culture shock. The mere notion that Eva and Eva* are neural duplicates is not by itself sufficient to make them phenomenal duplicates. To share an experience of phenomenal lack of attunement, they must be duplicate extended dynamic singularities.

To see this last point, imagine a scenario in which Eva and Eva* share identical neural states. They are synchronically identical in terms of the configuration of their brains. Eva, however, is living in her home country, Poland, and Eva* has left to take up a new life in Vancouver. Note that it would be Eva*, and not Eva, who experiences culture shock. What can we appeal to for an explanation of this difference in experience? The key difference is Eva's coupling to her local cultural environment. It is the coupling that is being intervened in in this scenario. Thus, it is Eva's relation to her cultural environment that makes the real difference in accounting for why Eva and Eva* could be neural duplicates and still differ in their phenomenal experience. Eva* experiences a lack of attunement with her cultural surroundings because the expectations that underlie her perception and action do not match those that are operative in her local cultural environment.

We take this thought experiment to show that pluggability fails at least for the case of phenomenal attunement. One cannot hold the internal states of the agent constant while varying the external states of the environments without this affecting whether a subject experiences phenomenal attunement. Eva is, in other words, nothing like Pac-Man. One cannot simply screen-off as background conditions, her ongoing coupling to her cultural environment since this coupling is constitutive of her conscious experience of phenomenal lack of attunement. The idea that Eva can be unplugged from her surroundings so long as her internal states are kept the same relies on the idea that Eva's experience is synchronically constituted. So long as you take two individuals who are internally the same at a given instant, this should necessitate that the individuals are also phenomenally identical. We take the failure of pluggability for the case of phenomenal attunement to follow from the diachronic constitution of phenomenal attunement. It is because phenomenal attunement is constituted by dynamical processes that interact over multiple timescales that individuals cannot simply be unplugged from one environment and plugged into another without this altering their experience of phenomenal attunement.

At this stage we anticipate some readers will raise the following worry: you state that it is not always possible to "unplug" the internal from the external without this having some non-trivial effect on phenomenal experience. But does the brain and its role in constituting consciousness not lend itself to this kind of "unplugging"? Think about cases such as dreaming, imagining, and mind wandering, in which the conscious mind is unplugged from the world, often allowing the brain to produce conscious states that are phenomenologically similar if not identical to those a subject enjoys when plugged into the world. We will call this the "internal fantasy objection" since it presses us to consider phenomenological similarities between perceptual experience and inner fantasies like dreaming and day-dreaming. If a subject can undergo a phenomenologically identical experience while being uncoupled from the world, doesn't this undermine our claim that coupling is what constitutes phenomenal attunement? Couldn't Eva dream she is back in Poland enjoying an experience

[9]The cognitive neuroscientist Georg Northoff (2019) would seem to agree. He argues that "what happens beyond the boundaries of our brain" in the body and in the surrounding environment is partially constitutive of consciousness (Northoff, 2019, p. 12). The brain's intrinsic dynamics should, he suggests, be thought of as "neuro-ecological" —that is to say, as "deeply embedded within and dependent upon the world" (Northoff, 2019, p. 7). We take this concept to be another way of talking about the unpluggability of the brain from the body and the rest of the world.

of phenomenal attunement with her surroundings, while she is actually asleep in her bed in Vancouver?

Again this objection assumes Eva's phenomenologically identical dream experience of attunement is the result of the brain states she undergoes at a moment in time. It assumes we can bracket Eva's history of coupling to her environment and consider only what is happening in her brain at the moment she is dreaming as constitutive of her experience, treating everything else as a background condition. We suggest that dreaming and waking experiences can be phenomenologically indistinguishable because the neural processes that are necessary for waking experience are recycled in sleep. In online perceptual experience internal (brain) and external (world) states are tightly coupled to one another via sensory and active states (Kirchhoff and Kiverstein, 2019a,b). Break the coupling and you break the possibility for the system in question to constitute conscious experience of phenomenal attunement. Online experience is the result of a dynamical coupling of perceiver and environment that unfolds over time. In fully offline, decoupled cases of experience, such as in dreaming or mind-wandering, internal and external states are not coupled in the same way. But offline cases of conscious experience, insofar as they recycle online perceptual experience, remain *indirectly* constitutively dependent on a history of coupling. Offline experiences inherit this constitutive dependence on coupling from online experience insofar as they are the result of recycling online experiences. Perhaps it will be objected that the indirect constitutive dependence of dream experience on coupling is really just a causal dependence. But again, this response assumes that we can take the neural processes that are constitutive of dream experience at a moment in time and bracket the longer history of coupling with the environment. We've been arguing that such an assumption is false at least for the case of phenomenal attunement.[10]

The ever persistent internalist skeptic will no doubt continue to insist on the intuition that Eva* can have the same phenomenal experience as Eva, whatever differences there might be between their respective environments. The modal intuition is familiar: once we fix the neural contribution to consciousness, variation in the environment of the individual is beside the point. The phenomenal experience of Eva and Eva* is fully metaphysically determined by whatever is taking place within their brains. Eva* could just as well be a disembodied brain floating about in space (Block, 2005). All that matters when it comes to her phenomenal experience is the configuration of neural activity in her brain. We will call this the "free-floating brain" objection. We do not pretend to know what would happen in such remote possible worlds in which disembodied brains can suddenly spring into existence. There may be, at the outer limits of this modal intuition, a possible world where Eva and Eva* could share the same phenomenal experience despite living in different environments. What is of interest to us are possible worlds closer to home. We therefore suggest that the modal intuition that stands behind the free-floating brain objection is quite beside the point when it comes to the case of culture shock.

[10]Our thanks to the reviewer for helpful discussion of this point.

WIDE AND DIACHRONIC CONSTITUTION: TWO CONCEPTUAL FLIPS

Internalist skeptics do not tire easily. We predict that they will continue to object, and most likely along quite familiar lines. It is something of a truism, they will insist, that cognitive activity (including conscious activity) is causally influenced by neural and non-neural (bodily, worldly) factors. But they will ask: How would we go about distinguishing non-neural bodily and worldly elements that are partially *constitutive* of the mind from such elements that merely *causal* influences on mental processes?

The causal-constitutive distinction has long dominated the debate about the extended mind (Adams and Aizawa, 2001, Adams and Aizawa, 2008; Rupert, 2004; Clark, 2008; Menary, 2010; Kirchhoff, 2015b; Kirchhoff and Kiverstein, 2019a). It will likely be objected that as defenders of ECM we are guilty of mistaking the causal dependence of phenomenal attunement on coupling with the cultural environment for the partial constitution of conscious experience by coupling with the cultural environment. Our opponents will assert that the proximal mechanisms internal to Eva's brain are minimally sufficient for her conscious experience. Let us stipulate that the existence of a population of neurons *N* is minimally sufficient for a conscious experience *C* if the activation of *N* is all that is required for the generation of *C*. Other neural activity may be causally necessary for the subject to come to instantiate *N*, but once the subject is in neural state *N*, no other neural activity in addition to *N* is required for the subject to experience *C* (Hohwy and Bayne, 2015, p. 159).[11] Our opponents will likely agree that the occurrence of *N* causally depends on a long prior history of engagement in cultural practices. Still they will say we must distinguish the proximate cause of Eva's experience in the here and now—the configuration of neural states *N* that constitute the minimal sufficient condition for Eva's experience—from whatever forms a part of the ultimate explanation for why Eva experiences what she does. It is only the proximate mechanisms that qualify as the realizers of her experience. The rest is a part of the ultimate explanation of why she has the experience she does. To insist otherwise is to commit the causal-constitutive fallacy.

This by now overly familiar line of argument is, in our view, premised on a number of problematic and mistaken assumptions. First, defenders of EM, most notably Clark and Chalmers (1998; but see also Wheeler, 2015), begin with the assumption that the basic ontological profile of the mind is a brainbound profile with the mind occasionally leaking out into the world. At least they assume that the brain plays a privileged role in constituting the mind. The paradigm of the mental is what goes on inside the head of individuals. The famous parity principle assesses

[11]Notice that this characterization of minimal neural sufficiency is very general. It is neutral on debates within the neuroscience of consciousness about the best theory of consciousness, such as the debate between the information integration theory of Tononi and colleagues, the global workspace theory of Baars and Dehaene, or the recurrent processing theory of Lamme and colleagues, to mention a few candidates. Our discussion need not take a stand on which of these theories is correct since we are concerned with the more general question of the correctness of the neural sufficiency claim.

the putative cognitive contribution of some external element by comparison with cognitive processes that take place internally inside of an individual's head.[12] This way of framing EM, however, concedes too much to the internalist, brainbound view of the mind. It assumes that the processes that take place inside of the brains of individuals are where constitutive causes are typically to be found. The external environment is populated with merely supporting causes, which may, under the right conditions, become constituent parts of a person's mental states. This is to accept that the environment can basically be screened-off from constitutive questions about the mind by processes that are internal to individuals.[13]

Our argument for ECM does not rest on such an internalist starting point. As we said at the outset of this paper, we consider that mentality is first and foremost constituted by bouts of temporally extended engagement with the environment. The vast majority of "what humans do and experience is best understood by appealing to dynamically unfolding, situated embodied interactions and engagements with worldly offerings" (Hutto et al., 2014, p. 1). One cannot uncouple the cognitive agent from its cultural, developmental, and historical environment because much of what the agent does constitutively depends on his or her taking part in cultural practices. Our internalist opponents claim that external elements can only play supportive causal roles, but they do so because they start from the assumption that minds are typically housed inside of the skin and skull of individuals and only occasionally if ever have recourse to go out into the world. This is an assumption that internalists ironically share with first-wave parity-based arguments for the extended mind. We, by contrast, take this assumption to be precisely what EM ought to challenge.

The EM debate has up until now largely played out around the question of how to delineate the boundaries of mind. Philosophers have wondered how to settle the question of where the mind stops and the rest of the non-mental world begins, and the debate has ended up being all about "location, location, (and only) location" (Di Paolo, 2009, p. 10). The argument about the boundaries of the mind is, however, not only about the spatial location of the mind, and whether the constituents or material realizers of a given class of mental states are sometimes wide or always narrow. We have been making an argument for EM on temporal grounds because we take the constitution of mind to be diachronic, not synchronic. The focus on location has led to a reification of the proximate-ultimate distinction. Once we think of mind as diachronically constituted, a strict choice between proximate and ultimate explanation is revealed to be a *false choice*. There are no fixed and sharp boundaries between proximate and ultimate causes. Cultural practices and biological processes are best conceptualized as elements of a single dynamical network (cf. Hurley, 1998; Kirchhoff and Kiverstein, 2019a).

Consider once more the case of Eva and Eva*. We argued that the main difference that determines why Eva* does, but Eva does not, experience culture shock is the coupling of the twins to the cultural environment. The constitution of Eva's conscious experience of culture shock is not wholly determined by her properties as a biological individual at a given snapshot moment in time. It is the dynamics of her coupling with her cultural environment that pick out the constituents that make up the minimally sufficient constitutive basis of her experience of phenomenal lack of attunement characteristic of culture shock. Perception and action couple Eva to her cultural environment, but this coupling unfolds over time. The coupling is in turn constrained by patterns of practice that also unfold over longer periods of time. It is the meeting up of these temporally extended processes that constitutes the conditions under which Eva, but not Eva*, experiences culture shock. So even when attempting to identify the minimally sufficient constitutive basis for certain kinds of conscious experience, one cannot simply separate the individual from her history. Constitution is not only wide; it is also *diachronic*.

The standard framing of the causal-constitutive distinction rests on a particular conception of the organism-environment relation; a conception according to which the world is "outside" or "external" to the organism and causes changes in its internal states. We have been arguing for ECM based upon the dynamics of the person's coupling with her cultural environment in perception and action. The person is situated within a larger dynamical process of the cultural practices she takes part in (Kirchhoff et al., 2018). Culture is not something external to the individual in which the individual is sometimes causally embedded. Both the individual agent and the cultural environment form out of a nesting of dynamical networks, including networks that form in the patterns of activities people engage in over long periods of time as members of cultural practices. The cultural environment is not outside of the individual. The individual is situated in a cultural environment in a way that calls into question any neat distinction between inside and outside.

Even if the reader agrees with us on all of these points, she might still raise the following objection: if cultural practices, unfolding over longer than synchronic timescales, are partly constitutive of conscious experience, then there is no stopping the rampant and out of control expansion of the mind into the world. This is the well known cognitive bloat objection to EM (cf. Sprevak, 2009; Rowlands, 2010). The arguments of this paper provide us with resources for replying to this worry.

Consider again our twin case: Eva and Eva* are in identical neural states, but Eva is living in her home country, Poland, while Eva* has left to take up her new life in Vancouver. Eva*, but not Eva, experiences culture shock. We have claimed the

[12]For example, in their discussion of Otto, Clark and Chalmers (1998) argue that the inscriptions in Otto's notebook are part and parcel of Otto's mind conditioned on an appeal to the function of brain-based biological memory. In a different example also considered by Clark and Chalmers (1998), the comparison is hypothetical, in the consideration of different instantiations of the function of a zoid-rotator in the game of Tetris. In both scenarios, however, it is the internal that is paraded as the benchmark for the mental. With this starting point in place, the real business of EM is to test whether specific external elements in the world play comparable or equivalent functional roles to those identified inside the head. If yes, then the external elements in question should fall within the confines of the mind.

[13]The notion of "screening-off" is standard in discussions of mental causation. As we pointed out in section "Synchronic and Diachronic Constitution," it is common procedure to look for causes that play the most salient role in the production of some phenomenon when identifying constituents of such a phenomenon.

key difference is coupling to the cultural environment. Our appeal to an account that makes a difference in determining and differentiating constitutive causes from mere background causes allows us to sidestep the cognitive bloat objection. We have argued that the phenomenology of culture shock can be formalized as the Kullback–Leibler divergence between prior expectations (P^*) and cultural practices (Po) generating sensory states, $C_{exp} = D_{KL}\ [P^* \,||\, P\text{o}]$, such that high misalignment (i.e., high uncertainty) between P^* and Po results in experiences of alienation and distress relative to current cultural practices. This leads to the following scenarios:

Poland: $C_{exp} = D_{KL}\ [P^* \,||\, P\text{o}] = 0$. Here Eva's expectations are aligned with her cultural world in such a fashion that she does not experience culture shock.

Vancouver: $C_{exp} = D_{KL}\ [P^* \,||\, P\text{o}] > 0$. Here Eva's expectations are misaligned with her cultural world in a way that results in her experiencing culture shock.

Counterfactually, were one to intervene in the cultural practices in Vancouver, one would expect a minimization in the divergence between P^* and Po, with a resulting change in Eva's phenomenology given the particular form of the agent-environment coupling. One might, for instance, point Eva to the district in Vancouver where a community of Polish emigres have made their home. Conversely, intervening in P^* would likely lead to similar results, a reduced sense of distress and alienation. This provides our argument for ECM with a methodology for identifying relevant (i.e., constitutive) causes, demarcating these from mere background causes such as oxygen in the atmosphere, given that the latter would at best be an indirect (i.e., background) cause of the generation of conscious experience. There is therefore a path by which to argue for ECM that does not lead to unconstrained spreading consciousness out into the world.

CONCLUSION

We have argued that the experience of phenomenal attunement is constituted by coupling to the cultural environment. A core structure of a person's conscious mental life is constituted by processes that criss-cross the boundary separating the brain from the body and the rest of the world. We've made such an argument based on the diachronic constitution of phenomenal attunement. Many hold that the proximate-ultimate distinction marks a sharp divide between causes that track *why* a system does what it does and *how* a system is able to do what it does. This distinction is taken by most to represent a division between diachronic (ultimate) and synchronic (proximate) explanation. We have argued that this choice between two different modes of explanation is a false choice. An explanation of phenomenal attunement needs to embed ultimate causes (cultural practices and histories of engagement with the world) within a proximate explanation of conscious experience. This has led us also to call into question the distinction between causation and constitution as it is generally deployed in the EM debate, taking steps toward a diachronic conception of constitution. Diachronic constitution implies that the agent and the wider cultural environment cannot be cleanly unplugged from one another in a way that would allow for a purely neural (synchronic) explanation of phenomenal attunement. Conscious persons cannot simply throw away the world and rely wholly on on-board neural resources for the generation of their conscious experience of being attuned to the world. Conscious beings cannot be unplugged from the extended dynamic singularity that forms in the agent's coupling with the world because conscious beings are extended dynamic singularities.

AUTHOR CONTRIBUTIONS

MK and JK have contributed equally to the production of this research article and approved the submitted version.

FUNDING

MK was funded by an Australian Research Council Discovery Project "Minds in Skilled Performance" (DP170102987). JK was funded as part of the European Research Council ERC Starting Grant (679190) awarded to Erik Rietveld.

REFERENCES

Adams, F., and Aizawa, K. (2001). The bounds of cognition. *Philos. Psychol.* 14, 43–64.

Adams, F., and Aizawa, A. (2008). *The Bounds of Cognition*. Hoboken, NJ: Wiley-Blackwell.

Anderson, M., Richardson, M. J., and Chemero, A. (2012). Eroding the boundaries of cognition: implications of embodiment. *Top. Cogn. Sci.* 4, 717–730. doi: 10.1111/j.1756-8765.2012.01211.x

Bennett, K. (2011). Construction Area (No Hard Hat Required). *Philos. Stud.* 154, 79–104. doi: 10.1007/s11098-011-9703-8

Block, N. (2005). Review of alva noë's action in perception. *J. Philos.* 102, 259–272. doi: 10.5840/jphil2005102524

Bruineberg, J., Kiverstein, J., and Rietveld, E. (2018). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese* 195, 2417–2444. doi: 10.1007/s11229-016-1239-1

Bruineberg, J., and Rietveld, E. (2014). Self-organisation, free energy minimisation and optimal grip on a field of affordances. *Front. Hum. Neurosci.* 8:599. doi: 10.3389/fnhum.2014.00599

Chalmers, D. (2008). *Foreword to Supersizing the Mind*. Oxford: Oxford University Press.

Chalmers, D. (2019). "Extended cognition and extended consciousness," in *Andy Clark and His Critics*, eds M. Colombo, L. Irvine, and M. Stapleton (Blackwell: Wiley).

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

Clark, A. (2009). Spreading the joy? Why the machinery of consciousness is (probably) still in the head. *Mind* 118, 963–993. doi: 10.1093/mind/fzp110

Clark, A. (2012). Dreaming the whole cat: generative models, predictive processing, and the enactivist conception of perceptual experience. *Mind* 121, 753–771. doi: 10.1093/mind/fzs106

Clark, A. (2016). *Surfing Uncertainty*. Oxford: Oxford University Press.

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 50, 7–19.

Cosmelli, D., and Thompson, E. (2010). "Embodiment or Envatment? Reflections on the bodily basis of consciousness," in *Enaction: Towards a New Paradigm in Cognitive Science*, eds J. Stewart, O. Gapenne, and E. Di Paolo (Cambridge, MA: MIT Press), 361–386.

Di Paolo, E. (2009). Extended Life. *Topoi* 28, 9–21.

Fletcher, P., and Frith, C. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536

Friston, K. (2018). Am I self-conscious? (Or does self-organisation entail self-consciousness?). *Front. Psychol.* 9:579. doi: 10.3389/fpsyg.2018.00579

Gallagher, S. (2018). *Enactivist Interventions: Rethinking the Mind*. Oxford: Oxford University Press.

Haugeland, J. (2002). "Andy Clark on cognition and representation," in *Philosophy of Mental Representation*, ed. H. Clapin (Oxford, UK: Oxford University Press), 24–36.

Hoffman, E. (1989). *Lost in Translation*. New York, NY: Penguin Books.

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Front. Psychol.* 3:96. doi: 10.3389/fpsyg.2012.00096

Hohwy, J. (2015). "Prediction error minimisation, mental and developmental disorder, and statistical theories of consciousness," in *Disturbed Consciousness: New Essays on Psychopathology and Theories of Consciousness*, ed. R. Gennaro (Cambridge, MA: MIT Press).

Hohwy, J., and Bayne, T. (2015). "The neural correlates of consciousness: Causes, confounds and constituents," in *The Constitution of Phenomenal Consciousness*, ed. S. Miller (Amsterdam, NE: John Benjamins).

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

Hurley, S. L. (2010). "The varieties of externalism," in *The Extended Mind*, ed. R. Menary (Cambridge, MA: MIT Press), 101–154.

Hutchins, E. (2011). Enculturating the supersized mind. *Philos. Stud.* 152, 437–446. doi: 10.1007/s11098-010-9599-8

Hutto, D. D., Gallagher, S., Ilundain-Agurruza, J., and Hipolito, I. (2019). "Culture in mind - an enactive account: Not cognitive penetration but cultural permeation," in *Schizophrenia and Common-Sense*, eds I. Hipolito, J. Gonçalves, and J. Pereira (Berlin: Springer).

Hutto, D. D., Kirchhoff, M. D., and Myin, E. (2014). Extensive enactivism: why keep it all in? *Front. Hum. Neurosci.* 8:706. doi: 10.3389/fnhum.2014.00706

Kim, J. (1990). "Supervenience as a philosophical concepts," in *Reprinted (1993) in Supervenience and Mind*, Ed. E. Sosa (Cambridge: Cambridge University Press), 131–160. doi: 10.1017/cbo9780511625220.009

Kim, J. (1998). *Mind in a Physical World*. Cambridge, MA: The MIT Press.

Kirchhoff, M. D. (2012). Extended cognition and fixed properties: step to a third-wave version of extended cognition. *Phenomenol. Cogn. Sci.* 11, 287–308. doi: 10.1007/s11097-011-9237-8

Kirchhoff, M. D. (2015a). Cognitive assembly: towards a diachronic notion of composition. *Phenomenol. Cogn. Sci.* 14, 33–53. doi: 10.1007/s11097-013-9338-7

Kirchhoff, M. D. (2015b). Extended cognition & the causal-constitutive fallacy: in search for a diachronic and dynamical conception of constitution. *Philos. Phenomenol. Res.* 90, 320–360.

Kirchhoff, M. D. (2015c). Species of realization and the free energy principle. *Aust. J. Philos.* 93, 706–723. doi: 10.1080/00048402.2014.992446

Kirchhoff, M. D., and Kiverstein, J. (2019a). *Extended Consciousness and Predictive Processing: A Third-Wave View*. Oxford: Routledge.

Kirchhoff, M. D., and Kiverstein, J. (2019b). How to demarcate the boundaries of mind: A Markov blanket proposal. *Synthese* doi: 10.1007/s11229-019-02370-y

Kirchhoff, M. D., Parr, T., Palacios, E., Friston, K., and Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *J. R. Soc. Interf.* 15:20170792. doi: 10.1098/rsif.2017.0792

Kiverstein, J. (2018). Free energy self: an ecological-enactive interpretation. *Topoi*.

Krueger, J., and Maiese, M. (2018). Mental institutions, habits of mind, and the extended approach to autism. *Thaumàzien* 6, 10–41.

Ladyman, J., and Ross, D. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Lamb, M., and Chemero, A. (2018). "Interacting in the open: where dynamical systems become extended and embodied," in *The Oxford Handbook of 4E Cognition*, eds A. Newen, L. De Bruin, and S. Gallagher (New York, NY: Oxford University Press).

Lawson, R. P., Rees, G., and Friston, K. (2014). An aberrant precision account of autism. *Front. Hum. Neurosci.* 8:302. doi: 10.3389/fnhum.2014.00302

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Basingstoke: Palgrave Macmillan.

Menary, R. (2010). *The Extended Mind*. Cambridge, MA: The MIT Press.

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

Nöe, A. (2006). Experience the world in time. *Analysis* 66, 26–32. doi: 10.1111/j.1467-8284.2006.00584.x

Noë, A. (2009). *Out of Our Heads: Why You are Not Your Brain, and Other Lessons From the Biology of Consciousness*. New York, NY: Hill and Wang.

Northoff, G. (2019). Lessons from astronomy and biology for the mind: copernican revolution in neuroscience. *Front. Hum. Neurosci.* 13:319. doi: 10.3389/fnhum.2019.00319

Palmer, C. J., Lawson, R. P., and Howly, J. (2017). Bayesian approaches to autism: towards volatility, action and behaviour. *Psychol. Bull.* 143, 521–542. doi: 10.1037/bul0000097

Pellicano, E., and Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* 16, 504–510. doi: 10.1016/j.tics.2012.08.009

Polger, T. (2010). Mechanisms and explanatory realization relations. *Synthese* 177, 193–212. doi: 10.1007/s11229-010-9841-0

Roepstorff, A., Niewöhner, C., and Beck, S. (2012). Enculturing brains through patterned practices. *Neural Netw.* 23, 1051–1059. doi: 10.1016/j.neunet.2010.08.002

Ross, D., and Ladyman, J. (2010). "The alleged coupling-constitution fallacy and the mature sciences," in *The Extended Mind*, ed. R. Menary (Cambridge,MA: The MIT Press), 155–166.

Rowlands, M. (2009). Enactivism and the extended mind. *Topoi* 28, 53–62. doi: 10.1007/s11245-008-9046-z

Rowlands, M. (2010). *The New Science of Mind*. Cambridge, MA: The MIT Press.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *J. Philos.* 9, 625–636.

Rupert, R. (2009). *Cognitive Systems and the Extended Mind*. New York, NY: Oxford University Press.

Sass, L. (1994). *The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic Mind*. Cornell, NY: Cornell University Press.

Schatzki, T. (1996). *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge, UK: Cambridge University Press.

Shapiro, A. (2011). *Embodied Cognition*. London: Routledge.

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

Slaby, J. (2016). Mind invasion: situated affectivity and the corporate life-hack. *Front. Psychol.* 7:266. doi: 10.3389/fpsyg.2016.00266

Spivey, M. (2007). *The Continuity of Mind*. Oxford: Oxford University Press.

Sprevak, M. (2009). Extended cognition and functionalism. *J. Philos.* 106, 503–527. doi: 10.5840/jphil2009106937

Sutton, J. (2010). "Exograms and interdisciplinarity: History, the extended mind, and the civilizing process," in *The Extended Mind*, ed. R. Menary (Cambridge, MA: The MIT Press), 189–225. doi: 10.7551/mitpress/9780262014038.003.0009

Sutton, J., Harris, C. B., Keil, P. G., and Barnier, A. J. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenol. Cogn. Sci.* 9, 521–560. doi: 10.1007/s11097-010-9182-y

van Gelder, T., and Port, R. (1995). "It's about time: An overview of the dynamical approach to cognition," in *Mind as Motion: Explorations in the Dynamics of Cognition*, eds R. Port and T. van Gelder (Cambridge, MA: The MIT Press), 1–44.

Ward, D. (2012). Enjoying the spread. Conscious externalism reconsidered. *Mind* 121, 731–751. doi: 10.1093/mind/fzs095

Wexler, B. (2008). *Brain and Culture: Neurobiology, Ideology, and Social Change*. Cambridge, MA: The MIT Press.

Wheeler, M. (2010). "In defense of extended functionalism," in *The Extended Mind*, ed. R. Menary (Cambridge, MA: The MIT Press), 245–270.

Wheeler, M. (2015). Not what it's like but where it's like: phenomenal consciousness, sensory substitution and the extended mind. *J. Conscious. Stud.* 22, 129–147.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 29 September 2020
doi: 10.3389/fpsyg.2020.546212

The Temporality of Situated Cognition

David H. V. Vogel[1,2], Mathis Jording[1,2], Christian Kupke[3] and Kai Vogeley[1,2]*

[1] Research Center Jülich, Institute of Neuroscience and Medicine (INM3), Jülich, Germany, [2] Department of Psychiatry, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany, [3] Department of Psychiatry, Society for Philosophy and Sciences of the Psyche, Charité, Humboldt-University Berlin, Berlin, Germany

Situated cognition embeds perceptions, thoughts, and behavior within the contextual framework of so-called "4E cognition" understanding cognition to be embodied, enactive, extended, and embedded. Whereas this definition is primarily based on the spatial properties of a situation, it neglects a fundamental constituent: the cognitive situation as *enduring*. On a subpersonal level, situated cognition requires the integration of information processing within a minimal temporal extension generating the basic building blocks of perception and action ("microlayer" of time). On a personal level, lived situations and experienced narratives leading to our biography can be defined by their broader temporal horizons ("macrolayer" of time). The macrolayer of time is based on and emerges from information processing on the microlayer of time. Whereas the constraints on the microlayer are primarily defined by the integrity of neurobiological processes within an individual cognitive system, the temporal horizons and subsequently the situational context on the macrolayer are defined by the complex affordances of a situation on a personal or interpersonal level. On both time layers, cognition can be defined as a continuous dynamic process, reflecting the transition from one situated state to another. Taken together, the events forming the delimiting horizons of these situations correspond to the temporal structure of the cognitive process along which it continuously proceeds. The dynamic driving and enabling this transition from state to state is synonymous with the inherent flow of time. Just as the layers of time, flow and structure, are inseparably connected. The integration of temporal flow and temporal structure into the continuous dynamic process constitutes the enduring situatedness of cognition. By providing everyday examples and examples from psychopathology, we highlight the benefits of understanding cognitive processes as part of enduring situations.

Keywords: temporality, time experience, dynamic systems, duration (time), cognitive processing, psychopathology

OPEN ACCESS

Edited by:
Achim Stephan,
University of Osnabrück, Germany

Reviewed by:
Lucia Maria Sacheli,
University of Milano-Bicocca, Italy
Maren Wehrle,
Erasmus University Rotterdam, Netherlands

****Correspondence:***
David H. V. Vogel
da.vogel@fz-Juelich.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *27 March 2020*
Accepted: *27 August 2020*
Published: *29 September 2020*

Citation:
Vogel DHV, Jording M, Kupke C and Vogeley K (2020) The Temporality of Situated Cognition. Front. Psychol. 11:546212. doi: 10.3389/fpsyg.2020.546212

INTRODUCTION

The proposal of situated cognition views cognitive processes as inherently located within a contextual framework comprising of a bodily (embodied), affording (enacted), physical (extended), and sociocultural (embedded) environment, also referred to as 4E cognition (Newen et al., 2018; Overmann and Malafouris, 2018). As the categories of 4E cognition are by definition intertwined

and inseparable from the processes they constitute, situated cognition attempts at examining, interpreting, and explaining the processes underlying cognition within situations defined by these categories. As an example, consider playing a musical instrument. The music is most likely influenced by the bodily state of the player such as, e.g., emotional states. In playing, these emotions as well as the production of the tones are embodied in the musician and automatically translated into bodily functions not primarily necessary for the act of playing (e.g., rocking movements, facial expressions; embodied). Further, the musician needs to be able to hear and evaluate their own playing of the instrument adaptively (e.g., the music is too loud because the keys are being hit too forcefully), or if playing in front of an audience may want to respond to the reactions of the audience and change the corresponding manipulation of the instrument accordingly (enacted). The musician may use an instrument to play and sheet music to remember a song (extended). And lastly, the musical genre and the instrument of choice may be influenced by social–cultural determinants and biographical influences (embedded). These four antecedents to situated cognition are often conceptualized as necessary preconditions to (human) cognition (e.g., Anderson, 2003; Niedenthal et al., 2005; Smith and Gasser, 2005; Bickhard, 2008; Gallagher, 2009). Although it is theoretically possible, cognitive processes, as they appear to us, do not happen without a physical form (embodied), are not entirely undirected (enacted), and do not appear without a physical (extended) or sociocultural (embedded) context. It has, however, been argued that aspects of a given cognitive act theoretically may be reduced to a point at which it is temporally unextended (Gallagher, 2000; Zahn et al., 2008; Arzy and Schacter, 2019). These conceptions run the danger of identifying the capacities of the cognizer purely by spatial properties. Cognition is subsequently located in a purely spatial *situs cogitans*. This site of thought can then be reidentified as the spatial relationships around or within it, or as the contents of the space surrounding and included in the *situs cogitans*. This reduction theoretically constrains all four aspects of situated cognition: embodiment is reduced to the concept of a cognitive process as placed in a location, which coincidentally is a body or a brain; enaction is minimized to a simple nonrecurring computation of inputs and outputs; extendedness refers to a merely spatial distribution of cognitive subfunctions and not explicitly to temporal extension; finally embeddedness may simply correspond to geometric relationships between the cognizer and their surroundings. Against the backdrop of cognition as situated 4E cognition, such an understanding of cognitive processes as timeless is hard to uphold and threatens the integrative understanding of cognitive processes. Reconsidering our example of playing a musical instrument, it is of course obvious that no aspect of it can be solely spatial. This leads to our proposal that temporality is necessarily a part of all aspects of a situated cognitive process. It is *embodied* in our lived experiences (Menary, 2008), *embedded* into our historical past (Kupke, 2009); it is it is *extended* in space, to be stored for later use or for us to be manipulated through present, future directed *enaction* (Gallagher et al., 2017).

As an alternative example, consider solving a simple math problem (Wilson and Clark, 2009). As an equation indicates a mathematical fact such as "$34 \times 12 = 408$," it can by itself be conceptualized as non-temporal. It could hence be argued that cognitive representation of the equation also may be situated within a non-temporal (purely spatial) *situs cogitans*. This would effectively imply recognizing the content of the equation as an objective fact standing outside of time. However, such a conception would not fit with how we usually interact with intellectual problems. Just as performing music, the cognition leading to the mathematical equation is a lived temporal act, too. The knowledge and abilities required to solve the equation depend on an enduring cognitive procedure with 4E properties. First, we (automatically) perform motor actions corresponding to the calculation, such as counting with our fingers, or experience bodily feelings such as frustration with related motor expressions in, e.g., facial expressions (*embodied*). Second, we type in individual numbers into the calculator or write down the equation in a specific way; in other words, we identify subaspects of the math problem and use them to manipulate our surroundings in order to solve the problem (*enacted*). Third, we may use aids such as pen and paper or a calculator to solve the problem (*extended*). Fourth and finally, the math problem solver is only presented with the problem and able to understand and deal with it because cognition is *embedded* into a sociocultural context. Despite our ability to potentially recognize, e.g., mathematical facts as outside of time and space, the cognitive act of performing the mathematical operations, as well as the cognitive acts of perceiving, understanding, and solving/interacting with such facts never are. In the following, we will lay out that any (cognitive) act must by default not only be conceptualized as extended, enacted, embodied, and embedded in space but also needs to be considered as enduring over time, necessitating temporality.

Temporality as the experience of and in time has long been identified as a key component of human existence and experience (e.g., Kant, 1781, p. 28ff; Husserl, 1928, p. 384ff). However, time is often spatialized (e.g., "time windows," "time distance," "time lines," etc.) and thus reduced to events and measures such as seconds or dates (Bergson, 1920, p. 82ff). Such conduct neglects the central role of temporality in our experience. As we pursue activities and carry out actions, time is ever present. While acting in the present, we enact control over our circumstances and wish to influence our (near or distant) future (Vogel et al., 2020). We do so by making use of past experiences and acquired knowledge. As such, our present can never be restricted to a "now" in the sense of an abstract point in time (Heidegger, 1953a, p. 406ff; Ricoeur, 1980). The present simultaneously includes a variety of temporal references and is necessarily a part of our ongoing narrative as an extended period of time. Time and temporality provide our actions with meaning and purpose regardless of their contents. Accordingly, any cognitive situation is not only spatially expanded but also temporally enduring. It is not merely a spatially expanded *situs cogitans*, but a temporally enduring situated cognition.

The research on social cognition has not been oblivious to temporality (e.g., Gallagher, 2009), and it is important

to recognize the implications of temporality of situated cognition, which would otherwise run the danger of being overlooked. Despite cognition implicitly and necessarily involving temporality, time is not of a trivial nature. Just as cognition is implicitly and necessarily embodied, enacted, extended, and embedded, it is temporally enduring. We believe that the characterization of cognition as a "dynamic process" (Clancey, 2008) accounts best for this specific feature of the temporality of situated cognition. To avoid the reduction of the cognizer to a *situs cogitans*, we will elaborate on these temporal principles of situated cognition and will illustrate its relevance for psychopathological disturbances. In order to avoid the misconception of the *situs cogitans*, we will first elaborate on the layers of time on which and through which cognition situates itself. Second, we will demonstrate how cognition is composed of a dynamic procedure, which determines the enduring situation. Furthermore, we will argue that the experience of time and duration reflect the abilities of the cognizer to successfully reach their goal. Finally, we will illustrate the relevance of temporality for psychopathological disturbances. To this end, we will attempt to draw from different theoretical approaches. We wish to demonstrate that despite differing and in some cases opposing ideas from these approaches (see Grush, 2006; Piekarski, 2017; Zahavi, 2018; for discussion), they may converge in important parts of their content and be put into a productive dialogue pushing the boundaries of current theory.

THE MACROLAYER AND THE MICROLAYER OF TIME

Time is of interest to us on two levels. We call the first phenomenally accessible level the macrolayer of time, or the biographic–personal time (Kupke, 2009, 2020; Vogel et al., 2020). The macrolayer describes time as it is open to conscious experience. This means both time as passing, for example, fast or slow, as well as our biographic narrative structuring past experiences and future plans. If we speak about situations on the macrolayer of time, we mean all temporal aspects consciously recognizable by a cognizer and of relevance to the currently ongoing cognitive process. This definition is not restricted to consciously, e.g., remembered or anticipated events, but refers to all potentially relevant aspects of an ongoing act, both explicitly and implicitly in the corresponding situation—both known and unknown to the cognizer independent from whether they have explicitly become aware of it or not.

Similar to the spatial aspects of a situation, the temporal features of a situation are defined by its borders, or "horizons" (Husserl, 1928, e.g., pp. 402, 411; Ricoeur, 1988; Gallagher et al., 2017). We argue that these borders define the present situation of a cognizer. For seemingly any action or activity, the horizons are foremost given by the reference points of beginning and end. Actions and steps performed in a situation are directed at the end of the situation starting from its beginning. At the horizons of a musical performance may, for example, stand the moments during which the first and the last notes were played. However, the temporal situation may extend beyond that specific period of the musical performance in the strict sense. In our example of the musician, depending on the context of the play and the context we as the observers with our own subjective point of view are interested in, the horizons of the situation may vary. For the musician's performance, the play may have started, e.g., when coming on stage for an audience, and it ended when it was left. Further, the performance may be part of an even wider situation, such as the musician's career opportunities and how they view themselves in relation to these opportunities.

Alternatively, we might not even be interested in the entire performance. We may want to observe the occurrence of a particular motif or theme in a piece of music. It then might be helpful to define horizons in terms of the movements of the musician to produce the motif or theme. By doing so, we identify a smaller duration within the subjective present. However, the way in which that particular motif is being played again may depend on its further temporal context, e.g., its position in the piece. Accordingly, at almost any graspable time scale, it would be possible to identify both larger and smaller time scales related to it. Although the observed individual need not be aware of this multitude of preconscious temporal horizons, they may all affect the execution of their play, as well as their observation.

This generalizes to any temporal situation. In the mathematical example, when the equation "$34 \times 12 = 408$" is presented to us on a piece of paper, the perception of the written numbers is not a singular event. In terms of an ongoing situated cognition, it is *embodied* in our lived experiences (Menary, 2008) as we have learned how to count and calculate in school; it is *extended* in space for us to be manipulated through future directed *enaction* (Gallagher et al., 2017) as we use our capacities to understand or validate the equation, and it is *embedded* into our historical past (Kupke, 2009) with math and numbers as historical derivatives. Hence, the situation is by default *enduring*. Importantly, these aspects of the cognitive act do not need to be explicitly known by the cognizer; they are, however, necessarily implicitly present in any cognitive act.

Theoretically, we are always able to pause, identify our present situation, and reflect upon it both in smaller and larger units. Human beings seem effortlessly able to perform this reflective task and position themselves and their actions in their individual time containing both a past and a future. Simultaneously, this potential overview on their own derived narrative (Ricoeur, 1980; Stanghellini and Mancini, 2017, p. 56f; Vogel et al., 2020) remains ever present, even while not consciously being aware of it. As we will elaborate in more detail below, the cognitive act is made possible by the constitution of the cognizer as continuously enduring along the structure of time.

These considerations of the potentially experienceable phenomenon hint at a related distinction between the *implicit* experience in time and the *explicit* experience of time. Time on the macrolayer is not always directly experienced, but remains in the background while we live our lives (Fuchs, 2005, 2013; Vogel et al., 2020). This implicit experience in time is observable in a variety of temporal phenomena, such as habit (Howell, 2015; Fuchs, 2018), corporeality/embodiment (Fuchs, 2005; Wehrle, 2020), historical circumstances (Kupke, 2009), and intersubjectivity/synchronicity (Fuchs, 2005, 2013;

Bloch et al., 2019) [also see von Gebsattel, 1954a, p. 137f for a similar distinction between experienced time ("erlebter Zeit") and lived time ("gelebter Zeit")]. However, this implicit temporality on the macrolayer of time is still potentially consciously accessible for us through either reflection or may impose on us under certain circumstances, as we will demonstrate in more detail in the section *Enduring Situatedness*.

When considering shorter temporal horizons, we notice that at some point of temporal reduction any smaller durations are no longer directly experienceable because the time interval that separates the corresponding horizons has become too short and is no longer open to conscious experience (Vogel et al., 2020). No later than then have we reached the microlayer or temporal–intentional layer of time (Vogeley and Kupke, 2007; Kupke, 2009). The microlayer of time is by definition not available to conscious experience but describes temporality as a necessary prerequisite to cognitive processing. To avoid confusion with the implicit temporality on the macrolayer of time, we will refer to the subpersonal microlayer processes as *intrinsic* temporality (Lenzo and Gallagher, 2020), although the terms sometimes have been used synonymously (Fuchs, 2013; Vogel and Vogeley, 2020).

The microlayer can be described by biological and (neuro-)physiological processes and phenomenological approaches. From a phenomenological perspective, the quintessential thoughts of Edmund Husserl are the most influential with respect to temporal consciousness (Husserl, 1928; Kupke, 2009). As an example, Varela (1999) has proposed that the diachronic unity of self-consciousness (Vogeley and Kupke, 2007) is achieved through the reciprocating oscillations of neural cell assemblies. Furthermore, Bayesian predictive processing recently has been used to account for the phenomenon of enduring consciousness (Hohwy et al., 2016; Wiese, 2017). Lastly, ongoing motor activity can be described in terms of motor cognitive models, such as, e.g., the forward model (Wolpert, 1997; Gallagher, 2000). In motor cognition, such models reflect the preconscious cognitive process of, e.g., performing a voluntary motor action.

Despite some of these theories and analyses being primarily directed at the concept of consciousness, whereas others concern cognition, they are all targeted at the same principle. More importantly, the underlying observation that temporal continuity is essential and inherent to the basic functioning of perception and experience is shared. Not only due to these similarities and convergences between the theories from different perspectives, we believe it essential to recognize that the operations on this level of time necessitate temporality. Because of its procedural nature, any such process is made up of consecutive, sometimes parallel and reemerging steps. These steps are taken by means of an inherent drive advancing and changing the content and condition of the cognitive process. As these changes occur over time, and because of their sequential and contingent nature, a dynamic process is not conceivable outside of time, but only in time.

As on the macrolayer, cognitive processes on the microlayer follow 4E properties including and in addition to their enduring quality. As taking place within a biophysical system, they are embodied; as goal-directed processes, they are enacted; as including peripheral and external information, they are extended and embedded. Finally, as continuously transitioning from one state to the next, relying on prior and expected states, the cognitive process on the microlayer is also to be characterized as enduring. In the math problem example, despite being able to appreciate the equation as a whole, we first need to appreciate every single digit, construct the numbers, understand the symbols, and derive the corresponding conclusions. Most of this process happens automatically, and only subsequently is the impression of the equation as a whole facilitated by the underlying microlayer processes. Importantly, any cognitively derived conclusion will need to undergo processing within the minimally enduring cognitive act.

Conclusively, this means that the horizon of each situation is contextually different both across acting individuals and observers. What may be defined as a present situation is highly dependent on the perspective of the cognizing and acting individual, as well as the observer's perspective. However, the overall form of a situation as a temporally extended and enduring (inter)subjective present that in turn is both composed of smaller temporal units and embedded into a larger biographical context cannot be ignored.

DYNAMIC PROCEDURES—THE FLOW AND STRUCTURE OF TIME

For any situation to be defined as temporal, it needs to be enduring and to take time. As we have argued in the previous section, temporal extendedness is a necessary condition for a cognitive process that claims to entail 4E properties. In order to integrate the multitude of perceptual information with prior knowledge and form a productive output, any cognitive mechanism needs time (Pöppel, 1997). Concerning the brain, this seeming triviality is first a biological and a physical property. In order to pass information from, e.g., the retina to the visual cortex a multitude of complex neurophysiological processes need to take place (Varela, 1999). Molecules need to change their configurations; action potentials have to be generated; transmitters are being released, traverse the synaptic cleft, and bind to receptors of the postsynaptic membrane of another neuron, from where new action potentials depart, etc. Just as the cognitive processes that these biochemical processes relate to, they are not only a localizable fact inside the neural system, but also follow temporal orders (Varela, 1999; Vogeley and Kupke, 2007). Not surprisingly, the time these processes necessarily take influences the temporal resolution of perception. Accordingly, as the neural processes take different amounts of time, the temporal resolution of different perceptual modalities differs too (Pöppel, 1997). Despite these perceptual limitations, our experiences appear to us as continuous (Husserl, 1928). Although it is open to debate whether the flow of consciousness is in fact continuous or only appears as such (e.g., Dainton, 2002; White, 2018), two observations seem obvious: (i) time moves toward the future and constitutes a passage; and (ii) time is organized along events that compose a structure (Kupke, 2009; Vogel et al., 2020).

It is important to keep in mind that despite the seemingly objective nature of these observations, with our approach we cannot make any deeper claim as to the nature of "objective time,"

be it biological, physical, or ontological in nature. The temporal situatedness as described herein primarily addresses temporality as it appears to subjective experience and cognition. Although we may believe the objective and the subjective to be intricately connected (Zahavi, 2018) when we propose that time appears as in motion, we mean that experiences are *felt* as moving into and toward the future. It seems impossible to stop time in its passage and transform experiences into a truly timeless nature. We experience our stream of perceptions and thoughts from a first-person point of view as being in a constant change of "pure transition" (Kupke and Vogeley, 2009, Kupke, 2020), and even states described as timeless have at least one thing in common with any other state: they end and turn into a different state.

If we try to explain cognitive states with the term "situated cognition" and apply this principle of ongoing experience, we reach the definition of the "dynamic process." The term "dynamic" or "transitional process" adequately describes the temporal properties of situatedness. Per definition processes are composed of stages of action, and in the case of the situated cognitive process, the cognizer cognizes through these stages. This process of cognizing necessarily implies a transition from one stage to the next to achieve an insight or a thought as the result of cognition. By virtue of the dynamic movement within the ongoing process, there is inevitable change in any such system. Additionally, it is not only the resulting cognitive state that is subject to change. Depending on the given affordances, the process itself may be altered in order to better address the context. Accordingly, we observe the dynamic in at least two forms of transition enabling continuous changes. One is the transition on the microlayer of time. Perceptions and impressions are replaced by the appearance of a following perception or impression. Integrative mechanisms within the brain receive and operate changing inputs. On a neurobiological level, action potentials and neural oscillations produce new events of the same type. The entire microlayer seems to follow the overall anisotropy of time itself (Vogel et al., 2020).

The second transition is observable on the macrolayer. When playing a piece of music, the sequential notes and the corresponding manipulations of the musical instrument need to be performed consecutively. The transition from one situation to the next is made up by a forward-directed movement that changes its content within one situational context (e.g., note after note within the piece of music). Accordingly, situations are composed of changing steps and in turn are themselves subject to change within the overarching biographical narrative.

Concerning time being structured, we notice that with the dynamic flow of time events necessarily follow one another. This pertains to the states of ongoing situations and therefore necessarily applies to their respective temporal horizons. A situation can only be properly delimited by an earlier and a later horizon if these horizons follow each other in a sensible order. We mostly structure our experiences along this temporal order of delineated meaningful events. It is further noteworthy that this structure is not restricted to past events but extends into the future by means of planning.

As with the passage of time, the structure appears both on the microlayer and the macrolayer of time (Kupke, 2009; Vogel et al., 2020). For the microlayer, Husserl's analysis of time consciousness (Husserl, 1928) examines the way in which our consciousness is continuously constituted by a process of retention, primal impression ("Urimpression"), and protention. Impressions describe the percept as appearing to/in consciousness at the border between retention and protention: retention entails the past impressions, and protention the coming impressions. Taken together, these structural components passively synthesize (Husserl, 1928) the appearance of an enduring continuous experience. Interestingly, Husserl proposes "horizons" to retention and protention (Husserl, 1928, e.g., pp. 402, 411) behind which the too-long past retentions slip, or too-far-ahead protentions are still hidden. We understand these horizons to effectively describe the borders of the smallest possible situation: the perceptual "now."

During any act encompassing this smallest structural entity, the microlayer of time facilitates the emergence of larger situations by executing their defining events. This macrolayer event structure visible in the multitude of larger horizontal situations has in its basics been depicted in the previous section. Effectively, the macrolayer's structure relates to the narrative of a situation (e.g., note A was played before note B, note C is being played now, and soon note D will be played before note E). At increasing time scales, the event structure becomes experientially increasingly coarse-grained, e.g., as a musical motif or theme, and turns into a person's account of a situation as a life phase (e.g., I was a student, became a teacher, and retired). This underlines the introductory assumption of the equation of the personal individual situation with the individual present. Unlike the moment of the "now," the "present" is identified as a consciously experienceable duration. It has repeatedly been argued that this "present" fluctuates along a duration of several seconds potentially corresponding to the time allocated to the integration of complex multimodal environments (for a recent discussion, see White, 2017). In other words, the minimal "present" appears to last several seconds.

As implied in the previous sections, the present situation is of a potentially variable extension interindividually, making the clear distinction difficult. This difficulty of identifying the present situation naturally stems from the relationship between structure and flow during the dynamic process. The continuous transition from "now" to "now" makes it implausible to identify the impeccable horizons of any situation. Pure transition causes the continuous emergence of a new situation with a new future horizon. Despite the obvious ability to prospectively and retrospectively identify singular events and their succession, the implicit flow of time smoothens over situations constituting a contiguous and meaningful whole.

As such, the procedural structure necessarily depends on the dynamic flow. The temporal flow first causes and enables our directedness toward the future horizon of situations. Hence, the concept of flow—at least on the macrolayer of time—may be understood as analogous to concepts such as "becoming" ("Werden") (Straus, 1928; von Gebsattel, 1954a,b; Fuchs, 2013) and "striving" ("Streben") (Minkowski, 1923, p. 220). Both these terms appositely recognize temporal flow as directed at something. In any situation, we are necessarily directed at

something. The directedness of temporal flow at the occurrence of specific desired, planned, or anticipated events emerges as a horizon of the situation. These horizons are then identifiable as the structure of time. Accordingly, this temporal structure is a structure the cognizer (en)actively defines only by means of their own temporal flow.

In the last section, we will explain how this intricate interrelation of flow and structure is met in the enduring situation. We will further provide everyday examples and examples from psychopathology to highlight the advantages of understanding cognition as enduring.

ENDURING SITUATEDNESS

What effectively remains missing in our observations is how the continuation of experience along events may account for the experience of duration in and of a given situation. In the following, we will demonstrate how the ever-present fusion of flow and structure into one dynamic process constitutes duration. It will become clear that the emergence and experienced variance of duration within a given situation are substantially influenced by and reflects a person's condition within his/her enduring situation.

In the last section, we saw that the term *dynamic process* appositely determines cognition as temporal flow and temporal structure. Flow and structure are not separable entities but determine each other (Vogel et al., 2020). The dynamic character engenders, coherently integrates, and joins the steps of the procedure. Simultaneously, the prospective and anticipated events of the procedure draw in the flow and give it direction. Although we may conceive a flow without specific direction, we do appreciate our own actions as directed toward something (Gallagher et al., 2017). This something consists of our environmental affordances (Gallagher, 2009; Gastelum, 2018). Again, these affordances are not thinkable without a dynamic and teleological directedness of the cognitive process toward the afforded events. Accordingly, what situations afford is already provided within their future horizons. We as individuals construct situations along these inherent potentialities. As the cognizer is a necessary part of the situation, the future horizon is not determined solely by objective conditions, but necessarily by the meaning the cognizer reaffords to the situation. Out from among all the potential prospects a situation may afford, the definite meaning is determined by the overarching temporal context as provided by the situated individual.

According to the work of Bergson (1920), this overarching temporal context is equitable with the subject as enduring. Our misconception of time as spatialized intervals and events, such as seconds, or days, hides our condition of being extended along an accumulation of these inseparable events, which in turn is the true *duration* of the individual. In our own words, the interconnected dependencies of flow and structure constitute the endurance of cognitive processes of an individual; the enduring situatedness of the cognizer is comprehensible as the maximal extension of the life situation: If we embed each situation into any hypothetical, larger situation, we eventually arrive at the "pure duration" (Bergson, 1920, p. 76f) as an overarching enduring of the cognitive process and the cognizer. This enduring process is defined and facilitated by the integration of the flow and the structure of time; in other words, the enduring process continuously flows within the structure it determines.

This integration is visible in the experience of the enduring situation itself. Importantly, though, while completely immersed in a given situation, we do not pay full attention to its enduring character; in other words, we do not notice time under usual conditions, and it remains implicit. This state has been referred to as "flow state" (Fuchs, 2005; Csikszentmihalyi et al., 2014). During such a state, the situated agent is undisturbed in their active becoming from the past toward the future, and the situation's future horizon is freely available. In other words, the situational affordance and the cognizer's reaffordance can be brought into agreement. It is the enduring situatedness as constituted by the interdependencies of structure and flow, which determines this agreement and allows time to remain implicitly in the background of experience. As stated above, this implicit experience in time is observable in a variety of different phenomena. However, time is consciously and explicitly experienced during a variety of different situations. These experiences and the underlying and relating cognitive processes can potentially be better understood by highlighting the temporality of situated cognition. While acting toward the future horizon in such situations, we become aware of our drive toward it. The most relatable experience may be the everyday phenomenon of boredom. A situation is boring, if the present cannot be brought into agreement with the horizon of the situated cognizer sufficiently. The person is unable to direct his/her action capacity at the desired goal. If, e.g., I wish to listen to an interesting musical performance, but the performance is not exciting enough, my overarching narrative horizons of listening to an interesting performance do not apply. Unfortunately, after I have already sat down in the audience, I cannot change the situation. If I am unable to leave the situation, I will inevitably feel bored, wishing to dedicate my time to something else (e.g., Fuchs, 2005; Elpidorou, 2018).

A second common experience is that of time pressure. The available time is known to be insuf?cient to perform an action directed at reaching a particular horizon and time appears as too little. Obviously, these two portrayals are only exemplary, and both cases do not describe all possible versions of time pressure and boredom [for a thorough analysis of boredom ("Langeweile"), see, e.g., Heidegger, 1953b]. Nevertheless, both the case of boredom and that of time pressure—and we argue any situation during which time is explicitly experienced—have one commonality in terms of the enduring situation: We recognize the limits of the situation and consciously experience its duration. The future horizon is either too close (time pressure) or too far away (boredom). In any case, the inner capacity to act toward a desired goal is impaired by the situation and consequently become aware of the duration of the situation. In other words, when the experience of duration imposes on us (Heidegger, 1953b), it corresponds to our inability to adequately reach a desired future horizon. Instead, we arrive too late or too early. The resulting question, as to the difference between the conceivable modes of time experience caused by these two conflicts between desired and imposed temporal horizons,

demonstrates that the broadened definition of the situated cognition as being *enduring* has significant implications for the cognizer and cannot be dismissed as trivial. As fully answering this question is well beyond the scope of this contribution, and it is not directly related to the general question of temporality in cognition as addressed herein, we wish to limit ourselves to the observation that the intricate interdependencies of flow and structure give rise to both the capacity of being active satisfactorily and its restrictions.

In this context of situatedness, satisfying activity relates to the opportunities available to an individual within time. Time is, as noted earlier, not a space where actions are counted as, e.g., seconds, but time is when we live and enact our enduring situation within its horizons. Our experience of being a situated agent acting in a dynamic reciprocity with our environment in space and during time accordingly describes the situatedness of the cognitive process.

ENDURING SITUATED COGNITION AND PSYCHOPATHOLOGY

An important implication of the temporality of cognitive processes lies in the study and understanding of psychopathological phenomena and mental disorders. As two examples, consider a psychotic episode in schizophrenia and a depressive episode in major depressive disorder (World Health Organization, 1993; American Psychiatric Association, 2013).

For patients with schizophrenia, disturbances in temporal processing have repeatedly been implicated (Fuchs, 2007; Vogeley and Kupke, 2007; Fletcher and Frith, 2009; Vogel D. et al., 2019). It has been proposed that a disruption or fragmentation in temporal continuity may lie at the heart of explaining psychotic symptoms. From a phenomenological point of view, it has primarily been argued that this fragmentation is due to an alteration of the future directed protention on the microlayer of time (Kupke, 2009, pp. 53–62; Fuchs, 2013; Stanghellini et al., 2016; Vogel D. H. V. et al., 2019; Vogel D. et al., 2019). In this context, protention is understood as analogous to an anticipatory process (Fuchs, 2013), which graduates future perceptions along a spectrum of probability. As protention fails in schizophrenia, new events cannot be sufficiently anticipated, leading to gaps in experience. These gaps in turn give rise to a variety of psychotic symptoms, including self-disorders, by means of an intrusion of thought or experience (Kupke, 2009, p. 55f; Fuchs, 2013; Stanghellini et al., 2016).

Relating this phenomenological view to current hypotheses from neuroscience, the cause for temporal fragmentation is hypothesized to lie in a faulty evaluation of predictions by the brain, caused by dysregulated dopamine release (Fletcher and Frith, 2009; Vogel D. H. V. et al., 2019). These inadequate predictions need to be explained by a dominant involvement of top-down processes that correlate to the development of delusional believes. The dysregulated dopaminergic activity causes the sudden emergence of aberrant affordances. In other words, otherwise meaningless objects and contexts suddenly appear as meaningful and important. This newly attributed relevance causes a directedness of the patient toward the aberrantly salient percept, which then has to be explained by reaffording them a personal meaning (i.e., delusion).

In terms of an enduring situated cognition, the situation of the cognizer during a psychotic episode is constantly disrupted by events unforeseeable and inexplicable to the cognizer. As we have seen, this idea of unpredictability is interestingly brought forth by both phenomenological approaches, as well as predictive processing hypotheses. In both cases, the structure of the cognitive procedure is fragmented primarily on the consciously inaccessible microlayer of time. However, the underlying dynamic driving the cognitive process remains intact. With still ongoing and enduring cognition the dynamic now connects the faulty procedural steps, forcing the cognizer to form beliefs and behaviors in accordance with an environment that is in constant and for the psychotic patient unpredictable change. This unpredictability hypothetically translates to any observation of the cognizer during psychosis, which effectively limits the transferability of individual behaviors to other persons with psychotic syndromes. The difference between the two approaches primarily lies in the formation of the symptom in question. Where, e.g., Fuchs (2013) suggests an intrusion of experiences into a "gap," which had already been caused by faulty protention, for predictive processing approaches the intrusion itself—caused by aberrant salience—is the disrupting/fragmenting factor.

As our second example, depressive episodes have been described by psychopathologists investigating the experience of time as being characterized by a "disturbance of becoming" or a "blocked future" (Straus, 1928; von Gebsattel, 1954a,b; Fuchs, 2013; Stanghellini et al., 2017; Vogel et al., 2018). Symptoms such as loss of interest, decreased affective reactivity, depressed mood, and emotionlessness, as well as psychomotor retardation, have been linked to this conceptual phenomenon (Stanghellini et al., 2017; Vogel et al., 2018). Patients during major depressive episodes are thought to have lost the biologically anchored strive toward the future, rendering them unable to form emotional connections and to pursue or even form goals (Straus, 1928). Patients lose the ability to act within the present, in order to change their environment and hence their future.

Putting these observations in terms of an enduring situated cognition, the depressive state can be understood as an unending situation. The dynamic of the cognitive process has changed to a point where the transition to a new situation no longer actualizes itself. The cognitive process is still generating an ongoing situation; however, this situation has lost key features of its temporal dynamic. The interaction between this change in the temporality of the situation in connection to the social and interpersonal embeddedness has been speculated to be a considerable part of patients' suffering by causing experiences of falling behind (Fuchs, 2013).

In both exemplary cases, these alterations in temporal experience have been described as hypothetically linked to changes in predictive processing (e.g., Kent et al., 2019; Vogel D. et al., 2019; Vogel D. H. V. et al., 2019; Vogel et al., 2020) just as temporality and predictive processing more generally have been (Grush, 2006; Hohwy et al., 2016; Wiese, 2017). Although these two approaches may not coincide in all respects [e.g., the debate between (neo-)neo-Kantian representationalism and Husserlian transcendental idealism (Zahavi, 2018), or

see Grush, 2006 for an overview of difficulties and weaknesses of some integrative approaches mentioned or referred to herein], they converge on the intrinsic unfolding of temporality on what we have now called the microlayer of time. Although these two examples from psychopathology require more elaboration, they demonstrate that information processing and its disturbances can be fruitfully reanalyzed and potentially be better understood by appreciating the enduring quality of cognition. Together with our examples concerning satisfying activity, they highlight the benefits of making explicit the temporality of cognitive acts. For future consideration, we propose that these changes of the temporal experience in altered mental states can be much better understood in terms of an altered enduring situatedness. Similar approaches may exist for other disorders and mental states for which temporality has been proposed as a defining constituent (e.g., Zukauskas et al., 2009; Hohwy et al., 2016; Vogel D. et al., 2019; Vogel D. H. V. et al., 2019; Vogel et al., 2020).

CONCLUSION

The enduring temporal context is constitutive of the situated character of cognition. With respect to the embodied nature, it contains the individual past in form of memories and schemata (Menary, 2008); as enacted and extended, it contains the individual future (Gallagher et al., 2017), as socioculturally embedded cognition contains the historical past (Kupke, 2009). We have argued that the time of the situated cognizer is lived both on a temporal–intentional microlayer reflecting a minimal duration of the cognitive process, and on a biographic–personal macrolayer reflecting an emerging narrative duration (Kupke, 2009; Vogel et al., 2020; also see Menary, 2008; Newen, 2018). We have further argued that situated cognition described as a dynamic process relates to the temporal properties of extended situations as being in flow, but at the same time following a structure. Lastly, we have demonstrated that the fusion of flow and structure engenders the capable and active agent and relates the experience of time to that of successful activity. It should now appear obvious that the premises of situated cognition overall describe cognition not in terms of a *situs cogitans* or site of thought, but as an enduring situation. Emphasizing the enduring quality of cognition is necessary to adequately describe the prerequisites to situated cognition, as well as the cognitive situation itself in order to foster a better understanding of cognitive acts in general, as well as during altered mental states.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

DV, CK, and KV provided the concept for the manuscript. DV wrote the first draft of the manuscript. All authors participated in the discussion, development, and proofreading of the manuscript. All authors approved the final version of the manuscript.

FUNDING

KV received funding from the Federal German Ministry of Education and Research (Grant No. 01GP1822, project consortium THERENIA) and from the EC, Horizon 2020 Framework Program, FET Proactive (Grant Agreement ID: 824128, project consortium VIRTUALTIMES).

REFERENCES

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5§)*. Washington: American Psychiatric Pub. doi: 10.1176/appi.books.9780890425596

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artif. Intell.* 149, 91–130. doi: 10.1016/s0004-3702(03)00054-7

Arzy, S., and Schacter, D. L. (2019). Self-agency and self-ownership in cognitive mapping. *Trend. Cogn. Sci.* 23, 476–487. doi: 10.1016/j.tics.2019.04.003

Bergson, H. (1920). *Sur les données imédiates de las conscience [Zeit und Freiheit]*. Germany: Frankfurt am Main.

Bickhard, M. H. (2008). *Is embodiment necessary? In Handbook of Cognitive Science*. Netherland: Elsevier, 27–40. doi: 10.1016/B978-0-08-046616-3.00002-5

Bloch, C., Falter-Wagner, C., Georgescu, A. L., and Vogeley, K. (2019). INTRApersonal Synchrony as Constituent of INTERpersonal Synchrony and its Relevance for Autism Spectrum Disorder. *Front. Robot. AI* 6:73. doi: 10.3389/frobt.2019.00073

Clancey, W. J. (2008). "Scientific antecedents of situated cognition," in *Cambridge Handbooks in Psychology*, Cambridge: Cambridge University Press.

Csikszentmihalyi, M., Abuhamdeh, S., and Nakamura, J. (2014). *Flow. In Flow and the foundations of positive psychology*. Netherland: Springer, 227–238. doi: 10.1007/978-94-017-9088-8_15

Dainton, B. (2002). *Stream of consciousness: Unity and continuity in conscious experience*. United Kingdom: Routledge. doi: 10.4324/9780203464571

Elpidorou, A. (2018). The good of boredom. *Philosoph. Psychol.* 31, 323–351. doi: 10.1080/09515089.2017.1346240

Fletcher, P. C., and Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536

Fuchs, T. (2005). Implicit and explicit temporality. *Philosoph. Psych. Psychol.* 12, 195–198. doi: 10.1353/ppp.2006.0004

Fuchs, T. (2007). The temporal structure of intentionality and its disturbance in schizophrenia. *Psychopathology* 40, 229–235. doi: 10.1159/000101365

Fuchs, T. (2013). Temporality and psychopathology. *Phenomenol. Cogn. Sci.* 12, 75–104. doi: 10.1007/s11097-010-9189-4

Fuchs, T. (2018). The cyclical time of the body and its relation to linear time. *J. Conscious. Stud.* 25, 47–65.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trend. Cogn. Sci.* 4, 14–21. doi: 10.1016/s1364-6613(99)01417-5

Gallagher, S. (2009). "Philosophical antecedents to situated cognition," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins., and M. Aydede (United States: Cambridge University Press), 35–51. doi: 10.1017/cbo9780511816826.003

Gallagher, S., Martínez, S. F., and Gastelum, M. (2017). "Action-space and time: Towards an enactive hermeneutics, in *Place, space and hermeneutics*. B. Janz (Cham: Springer), 83–96. doi: 10.1007/978-3-319-52214-2_7

Gastelum, M. (2018). Temporal experience from a 4E perspective. *Adapt. Behav.* 26, 269–272. doi: 10.1177/1059712318790752

Grush, R. (2006). How to, and how not to, bridge computational cognitive neuroscience and Husserlian phenomenology of time consciousness. *Synthese* 153, 417–450. doi: 10.1007/s11229-006-9100-6

Heidegger, M. (1953a). *Sein und Zeit*. Germany: Max Niemeyer Verlag.

Heidegger, M. (1953b). "Die Grundbegriffe der Metaphysik," In *Klostermann Frankfurt*, ed, F. W. Von Herrmann, (Germany: Klostermann Frankfurt).

Hohwy, J., Paton, B., and Palmer, C. (2016). Distrusting the present. *Phenomenol. Cogn. Sci.* 15, 315–335. doi: 10.1007/s11097-015-9439-6

Howell, W. (2015). Learning and the Development of Meaning: Husserl and Merleau-Ponty on the Temporality of Perception and Habit. *South. J. Phil.* 53, 311–337. doi: 10.1111/sjp.12116

Husserl, E. (1928). "Vorlesungen zur Phänomenologie des inneren Zeitbewußtseins," in *Max Niemeyer Verlag Tübingen*, M. Heidegge, (Berlin: Springer).

Kant, I. (1781). *Critique of Pure Reason*. New York: Prometheus books.

Kent, L., van Doorn, G., Hohwy, J., and Klein, B. (2019). Bayes, time perception, and relativity: The central role of hopelessness. *Conscious. Cogn.* 69, 70–80. doi: 10.1016/j.concog.2019.01.012

Kupke, C. (2009). *Der Begriff Zeit in der Psychopathologie*. Berlin: Parodos Verlag.

Kupke, C. (2020). "Zeiterleben als Erleben von Zeit Ein philosophischer Versuch," in *Palliativ und Zeiterleben*, eds H. Ewald, K. Vogeley, and R. Voltz (Germany: Kohlhammer), 21–35.

Kupke, C., and Vogeley, K. (2009). "Constitution of cognition in time," in *Chronobiology and Chronopsychology*, eds T. G. Baudson, A. Seemüller, and M. Dresler (Germany: Pabst Science Publishers), 121–149.

Lenzo, E., and Gallagher, S. (2020). "Intrinsic Temporality in Depression: Classical phenomenological psychiatry, affectivity and narrative," in *Time and Body: Phenomenological and Psychopathological Approaches*, eds C. Tewes and G. Stanghellini (Cambridge, UK: Cambridge University).

Menary, R. (2008). Embodied narratives. *J. Consc. Stud.* 15, 63–84.

Minkowski, E. (1923). Bleulers schizoidie und syntonie und das zeiterlebnis. *Z. Für Gesamte Neurol. Psychiatr.* 82, 212–230. doi: 10.1007/bf02970889

Newen, A. (2018). The embodied self, the pattern theory of self, and the predictive mind. *Front. Psychol.* 9:2270. doi: 10.3389/fpsyg.2018.02270

Newen, A., De Bruin, L., and Gallagher, S. (2018). *The Oxford handbook of 4E cognition*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780198735410.001.0001

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., and Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personal. Soc. Psychol. Rev.* 9, 184–211. doi: 10.1207/s15327957pspr0903_1

Overmann, K. A., and Malafouris, L. (2018). *Situated Cognition*. Netherland: Springer, 1–8. doi: 10.1002/9781118924396.wbiea2201

Piekarski, M. (2017). Commentary: Brain, Mind, World: Predictive Coding, Neo-Kantianism, and Transcendental Idealism. *Front. Psychol.* 8:2077. doi: 10.3389/fpsyg.2017.02077

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trend Cogn. Sci.* 1, 56–61. doi: 10.1016/S1364-6613(97)01008-5

Ricoeur, P. (1980). Narrative time. *Crit. Inqu.* 7, 169–190. doi: 10.1086/448093

Ricoeur, P. (1988). *Time and Narrative*. Chicago: University Press of Chicago.

Smith, L., and Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artif. Life* 11, 13–29. doi: 10.1162/1064546053278973

Stanghellini, G., Ballerini, M., Presenza, S., Mancini, M., Northoff, G., and Cutting, J. (2017). Abnormal time experiences in major depression: an empirical qualitative study. *Psychopathology* 50, 125–140. doi: 10.1159/000452892

Stanghellini, G., Ballerini, M., Presenza, S., Mancini, M., Raballo, A., Blasi, S., et al. (2016). Psychopathology of lived time: abnormal time experience in persons with schizophrenia. *Schizophr. Bull.* 42, 45–55. doi: 10.1093/schbul/sbv052

Stanghellini, G., and Mancini, M. (2017). *The therapeutic interview. Emotions, values, and the life-world*. Cambridge: Cambridge University Press. doi: 10.1017/9781316181973

Straus, E. W. (1928). Das zeiterlebnis in der endogenen depression und in der psychopathischen verstimmung. *Monatsschr. Psychiat. Neurol.* 68, 640–656. doi: 10.1159/000164543

Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. *Natur. Phenomenol. Iss. Contemp. Phenomenol. Cogn. sci.* 64, 266–329.

Vogel, D. H., Falter-Wagner, C. M., Schoofs, T., Krämer, K., Kupke, C., and Vogeley, K. (2020). Flow and structure of time experience–concept, empirical validation and implications for psychopathology. *Phenomenol. Cogn. Sci.* 19, 235–258. doi: 10.1007/s11097-018-9573-z

Vogel, D. H., Krämer, K., Schoofs, T., Kupke, C., and Vogeley, K. (2018). Disturbed experience of time in depression—evidence from content analysis. *Front. Hum. Neurosci.* 12:66. doi: 10.3389/fnhum.2018.00066

Vogel, D. H. V., Beeker, T., Haidl, T., Kupke, C., Heinze, M., and Vogeley, K. (2019). Disturbed time experience during and after psychosis. *Schizophr. Res. Cogn.* 17:100136. doi: 10.1016/j.scog.2019.100136

Vogel, D., Falter-Wagner, C. M., Schoofs, T., Krämer, K., Kupke, C., and Vogeley, K. (2019). Interrupted time experience in autism spectrum disorder: empirical evidence from content analysis. *J. Autism Devel. Disor.* 49, 22–33. doi: 10.1007/s10803-018-3771-y

Vogel, D. H. V., and Vogeley, K. (2020). "Time Experience in Autism Spectrum Disorder," in *Encyclopedia of Autism Spectrum Disorders*, ed. F. Volkmar (New York, NY: Springer). doi: 10.1007/978-1-4614-6435-8_102354-1

Vogeley, K., and Kupke, C. (2007). Disturbances of time consciousness from a phenomenological and a neuroscientific perspective. *Schizophr. Bull.* 33, 157–165. doi: 10.1093/schbul/sbl056

von Gebsattel, V. F. (1954a). *Die Störungen des Werdens und des Zeiterlebens im Rahmen psychiatrischer Erkrankungen. In Prolegomena einer medizinischen Anthropologie*. Berlin: Springer, 128–144. doi: 10.1007/978-3-642-87964-7_5

von Gebsattel, V. F. (1954b). *Zeitbezogenes Zwangsdenken in der Melancholie. In Prolegomena einer medizinischen anthropologie*. Berlin: Springer, 1–18. doi: 10.1007/978-3-642-87964-7_1

Wehrle, M. (2020). Being a body and having a body. *The twofold temporality of embodied intentionality**. *Phenomenol. Cogn. Sci.* 51, 128–33. doi: 10.1007/s11097-019-09610-z

White, P. A. (2017). The three-second "subjective present": A critical review and a new proposal. *Psychol. Bull.* 143:735. doi: 10.1037/bul0000104

White, P. A. (2018). Is conscious perception a series of discrete temporal frames? *Consc. Cogn.* 60, 98–126. doi: 10.1016/j.concog.2018.02.012

Wiese, W. (2017). "Predictive Processing and the Phenomenology of Time Consciousness - A Hierarchical Extension of Rick Grush's Trajectory Estimation Model," in *Philosophy and Predictive Processing: 26*, eds T. Metzinger and W. Wiese (Frankfurt am Main: MIND Group). doi: 10.7551/mitpress/9780262036993.003.0008

Wilson, R. A., and Clark, A. (2009). "How to Situate Cognition: Letting Nature Take its Course," in *The Cambridge Handbook of Situated Cognition*, eds M. Aydede and P. Robbins (Cambridge: Cambridge University Press), 55–77. doi: 10.1017/CBO9780511816826.004

Wolpert, D. M. (1997). Computational approaches to motor control. *Trend. Cogn. Sci.* 1, 209–216. doi: 10.1016/S1364-6613(97)01070-X

World Health Organization (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. Switzerland: World Health Organization.

Zahavi, D. (2018). Brain, Mind, World: Predictive coding, neo-Kantianism, and transcendental idealism. *Husserl. Stud.* 34, 47–61. doi: 10.1007/s10743-017-9218-z

Zahn, R., Talazko, J., and Ebert, D. (2008). Loss of the sense of self-ownership for perceptions of objects in a case of right inferior temporal, parieto-occipital and precentral hypometabolism. *Psychopathology* 41, 397–402. doi: 10.1159/000158228

Zukauskas, P. R., Assumpção, F. B. Jr., and Silton, N. (2009). Temporality and Asperger's syndrome. *J. Phenomenol. Psychol.* 40, 85–106. doi: 10.1163/156916209X427990

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 December 2020
doi: 10.3389/fpsyg.2020.607917

An Algorithmic Metaphysics of Self-Patterns

*Majid D. Beni**

Department of Philosophy, Middle East Technical University, Ankara, Turkey

The paper draws on an algorithmic criterion to demonstrate that the self (as initially described in Shaun Gallagher's a pattern theory of self) is a composite, scattered, and patterned object. It also addresses the question of extendedness of the self-pattern. Based on the criteria drawn from algorithmic complexity, I argue that although the self-pattern possesses a genuinely extended aspect (and in this sense, the self-pattern is minimally extended) the self-pattern and its environment do not constitute a genuine composite object.

Keywords: coupling, constitution, composition, compressibility, free energy principle, self-pattern

INTRODUCTION[1]

This paper does not intend to present a general theory of extendedness. Instead, it focuses on a specific case to address the question of *the relations between the self and its environment* at a fundamental level. The paper endorses the criterion of compressibility-cum-simplicity as the yardstick of the extendedness of *the self*. The general intuition behind this move from the theory of extended cognition to the discussion of extendedness of the self is this: the self is a cognitive agent *par excellence*, and if we unravel the issue of the extendedness of the self adequately (in terms of complexity and simplicity of *patterns*) we will acquire deep insights into the criterion of extendedness of cognition. I conceive of the relationship between the self and its extension in terms of Gallagher's (2013) "A Pattern Theory of Self" (also see Kyselo, 2014; Beni, 2016; Gallagher and Daly's, 2018).

Although Gallagher and colleagues speak extensively about the dynamical relations between aspects of the self-pattern (the extended aspect included), they do not address the issue of the ontological state of the patterned self and its aspects. The only exception is their expressed sympathy for Dennett's (1991) theory of real patterns (Gallagher and Daly's, 2018, p. 2). The paper considers the self as a composite object and asks two important questions;

1. Is the extended aspect constitutive of the self(-pattern)? If that is the case, there is some purchase for the extendedness of the self under the pattern theory.
2. Do the self-pattern and the environment constitute a genuine composite object?

The notion of "constitution" that is used in the present paper is different from Gallagher's view of "constitution." This paper regards "constitution" as a matter of composition, whereas Gallagher (2018) speaks of "dynamical constitution." "Dynamical constitution" is a term of art, and it could be (and indeed has been) explicated in terms of reciprocal (or circular) causality. Others have defended this view before. Kirchhoff (2015), for example, identifies "constitution" with

[1]The paper has benefited a lot from the contribution of two referees of this journal, the editor, and steve elliot, these debts are gratefully acknowledged.

OPEN ACCESS

Edited by:
Albert Newen,
Ruhr University Bochum, Germany

Reviewed by:
Mateusz Wozniak,
Central European University, Hungary
Shaun Gallagher,
University of Memphis, United States

***Correspondence:**
Majid D. Beni
mbeni@metu.edu.tr;
davoody1980@gmail.com

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *18 September 2020*
Accepted: *13 November 2020*
Published: *04 December 2020*

Citation:
Beni MD (2020) An Algorithmic Metaphysics of Self-Patterns. Front. Psychol. 11:607917. doi: 10.3389/fpsyg.2020.607917

diachronic causal coupling. Kirchhoff's view is in harmony with the extended-enactivist approach (as well as with Gallagher's use of dynamical gestalt). However, I have some reservations about how to construe the "causal coupling" relation in more or less familiar metaphysical terms. I shall unpack this reservation in the remainder of the paper, but for the time being suffice to say that the concern about the causal-coupling notion of constitution is discussed under the coupling-constitutive fallacy (Aizawa, 2010). The fallacy holds that the causal coupling relation is not sufficient for the constitution. And although Gallagher does address the causal-constitution fallacy (Gallagher, 2018), in agreement with the enactivist approach (Kirchhoff, 2015), he eventually renounces the compositional view on "constitution" and eradicates the difference between the notion of "constitution" and "causality" and argues that "dynamical couplings of brain-body-environment constitute the mind" (Gallagher, 2018, p. 208). As I say, I do not engage in a fundamental debate about the plausibility of enactivism. Nor do I claim that Gallagher's approach simply begs the question of extendedness of the mind. I just claim that his dedication to enactivism and extendedness are not well supported enough. That is to say, despite the remarkable success of enactivism and extendedness in making sense of psychological theories (Varela et al., 1991; Barsalou, 2008; Pezzulo et al., 2012; Bitbol and Gallagher, 2018), these approaches have not been developed into a well-supported metaphysical stance, say, about the objecthood or reality of the extended objects. This does not need to mean that the dynamical (causal-coupling) conception of constitution is generally wanting. Indeed, Gallagher does provide a rather detailed critical discussion of the new mechanist roots of the constitution-coupling notion of constitution (Bitbol and Gallagher, 2018; Gallagher, 2018)[2]. But the problem is that this approach does not expansively elaborate on the ontological aspects of the extended objects. I take a compositional stance on the question of constitution. In defense of this move, I can say that the compositional stance could be developed into a clear metaphysical interpretation of real patterns as well as self-patterns. At the same time, this proposal is unassuming, in the sense that it does not intend to deny the viability of dynamical approaches. Nor does it claim the ultimate superiority of the compositional approach.

Perhaps it was wise, on Gallagher's part, to take enactivism as a basic perspective whose soundness does not need to be supported by further philosophical argument. But I assume that the compositional view on the constitution (that gives rise to the notorious coupling-constitution fallacy) is as respectable as the rival stance. The present paper pays allegiances to the classical compositional view. It could be granted that a compositional view on the constitution of the self and its relationship with its environment deserves to be heard out too. I pinpoint the need for a compositional approach with a reference to the coupling-constitutive fallacy before going further.

The question of the constitutive/coupling relation between the self and the environment is dictated with an eye to a serious challenge to the theory of extended cognition (Clark and Chalmers, 1998); for the challenge, see Aizawa, 2010; Adams and Aizawa, 2012). I suggest that the criteria of compressibility and simplicity of patterns provide precious insight into the question of extendedness of the self. The proposed criteria are inspired by Steve Petersen's (Petersen, 2013, 2019) algorithmic metaphysics of the composition of objects (and more originally, by Dennett's (1991) theory of real patterns, as well as Ladyman and Ross's (2007) engagement with the idea of real patterns). Dabbling in algorithmic metaphysics in this fashion, I provide answers to the two abovementioned questions. I argue that;

- In answer to (1) above, the extended aspect is constitutive of the self-pattern. The answer is backed up by the criterion of compressibility and its minimal and maximal clauses (as inspired by Petersen's work). I submit that the self-pattern is a composite object constituted by various aspects.
- In answer to (2) above, I suggest that the self-pattern and the environment do not constitute a composite object. I substantiate this point by invoking the same criteria of compressibility and simplicity. I draw on Friston et al.'s theory of selfhood under the Free Energy Principle to present the criteria of compressibility and simplicity of the self to substantiate my claim.

The paper is structured in the following way. I use a broad brush to sketch some platitudes about the extended cognition thesis as well as the coupling-constitution fallacy. Then I focus on Gallagher's *a pattern theory* of the self and expose the question of the relationship between various contributors to the self (the cognitive aspect and the extended aspect included). Then I outline Petersen's patternist criterion of being a composite object and show that the self-pattern is a scattered composite object that subsumes various aspects, elements, and factors as its constituents (the extended aspect included). This provides some purchase for defending a minimal account of the extendedness of the self-pattern (only in the sense that the self has an extended aspect). Then I show that the self-pattern and the environment do not constitute a composite pattern (in an ontologically significant way). The conclusion is that although the self-pattern is minimally extended, the extendedness only amounts to a coupling relation between the self-pattern and its environment but not to a genuine form of constitution.

THE EXTENDED COGNITION AND THE COUPLING-CONSTITUTION FALLACY

According to the extended cognition thesis, to fulfill our cognitive goals, we depend on elements in our environment, including the "technological gadgets with which we regularly and uncritically interact" (Carter and Kallestrup, 2019, p. 1). The insight into the integration between the cognitive abilities of the organic agents and (presumably) non-organic (or extra-organismic) devices led to the philosophical belief that the extra-organismic devices constitute a non-negligible part of the cognitive process (Clark and Chalmers, 1998). The environmental factors are not only coupled with cognitive processes, they also constitute

[2]I owe this important remark to one of the reviewers of this paper.

parts of these cognitive processes. According to Clark and Chalmers (1998, p. 8):

> 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.

Despite its natural appeal, the extended mind thesis has been targeted by the objection from the coupling-constitution fallacy (Adams and Aizawa, 2008, 2009; Aizawa, 2010). The objection is based on a denial: all external resources causally interact with cognitive systems are not integrated with cognitive systems.

The main argument behind the extended mind thesis seems to be something like this: Assume that X is a cognitive process. X is causally dependent on Y and is mutually interacting with it to fulfill its cognitive goals (meaning that Y is not dangling at the end of the causal chain and is included in the loop). It follows that X and Y form an integrated cognitive process. The argument is allegedly based on some fallacy: from the fact that X and Y are causally connected, it does not follow that X and Y form an integrated cognitive process, given that causation simpliciter is not enough for supporting the claim about the constitution of the system (Aizawa, 2010, p. 333). One reason for this pessimism is that "constitution" is considered to be synchronic whereas causality needs to be diachronic. The pessimism could as well be based on some deeper skepticism about the capacity of "causality" to be the universal glue of constitution [As it happens, skepticism about the status of causality finds its way into the work of some notable pattern theorists (Ladyman and Ross, 2007, chapter 5)]. Be that as may, according to Adams and Aizawa (2008, p. 91) "It simply does not follow from the fact that process X is in some way causally connected to a cognitive process that X is thereby part of that cognitive process."

There are various sorts of reactions to the coupling-constitution fallacy (Wilson, 2004, 2010; Clark, 2008; Rowlands, 2009; Walter and Kyselo, 2009; Piredda, 2017). Advocating an enactivist approach, Gallagher himself considered the coupling-constitution fallacy shortly and let it down rather easily. He asserts that the "diachronic conception of constitution that includes reciprocal causal relations" can be adopted by the enactivist, extended-mind approach, which supports a dynamical and holistic conception of cognition (Gallagher, 2018, p. 207). It is by no means a shortcoming of Gallagher's approach that it depends on a causal-coupling conception of constitution. This dynamical conception is well-supported enough (Kirchhoff, 2015; Kirchhoff and Kiverstein, 2019). So, instead of trading off intuitions about the (undeniable) appeal of enactivism, I suggest a rather operational (i.e., algorithmic) criterion of testing the metaphysical viability of extendedness. In this respect, the endeavor of this paper is different from the preceding debates of critics and advocates of enactivism and the extended mind theory who do not provide a clear demarcation criterion for distinguishing genuine cases of constitution from cases of coupling simpliciter. Let me elaborate.

The question of extendedness is this: where to draw the boundaries of cognition? The thesis of extendedness indicates that there are no sharp boundaries between the cognitive system and its environment, and the cognitive system and its environment constitute an entity. I submit that the philosophy of selfhood provides a good framework for unraveling the question of extendedness. The self is the cognitive agent par excellence, and the question of extendedness could be rephrased in terms of how to draw the boundaries that separate the self-pattern from its environment. The discussion will be continued in the next section.

A PATTERN THEORY OF SELF

I address the question of extendedness of the self in the context of Gallagher's (2013) a pattern theory of self (I call it *the* pattern theory but without any specific philosophical intentions). The pattern theory of the self has been discussed expansively (Kyselo, 2014; Newen, 2018; Beni, 2019a,b). While the choice of "the pattern theory" in the context of the present enquiry is to some extent arbitrary, the theory provides a nice venue for pursuing the question of extendedness. This is because the pattern theory stipulates the existence of an extended aspect of the self.

Gallagher states the pattern theory of self in the following manner: "According to the pattern theory, a self is constituted by some characteristic features or aspects that may include minimal embodied, minimal experiential, affective, intersubjective, psychological/cognitive, narrative, *extended*, and *situated* aspects" (Gallagher, 2013, p. 1 emphasis added). According to this approach, what is called the self is a cluster concept that includes a sufficient number of features. Despite speaking of *aspects* of the self, Gallagher endeavors to "stay plural about the concept of self" (Gallagher, 2013, p. 1). If so, the so-called aspects (as being organized into certain patterns according to Gallagher) are not models of something (i.e., the self) that has its independent existence. The point about the existence is rather important in the context of our paper (which is concerned with *metaphysical* issues).

The self is not a simple entity with its independent existence. However, it is not obvious that the self-pattern does not exist at all (I will follow Gallagher, 2013 and use "self-pattern" and "self" interchangeably). The pattern theory does not advocate a form of eliminativism about the self (Metzinger, 2003). The self-pattern is not non-existent in the context of the theory. What manner of existence does the self-pattern possess then? From the metaphysical point of view, we can assume that the self-pattern (which is neither independently existent nor totally non-existent) exists as a composite object, constituted by the menagerie of various aspects and elements. The self-pattern and its aspects and elements do not exist independently of one another. In this paper, I argue that the elements and aspects of the self are constitutive of the self-pattern, which is a scattered composite object.

To a first approximation, Gallagher's definition of "self-pattern" does not provide a clear insight into the ontological

status of the self. Gallagher suggests that "what we call self consists of a complex and sufficient pattern of certain contributors, none of which on their own is necessary or essential to any particular self" (Gallagher, 2013, p. 3). What is the relation between contributors of the self? The pattern theory emphasizes the diversity of aspects and elements of the self. However, it does not account for the relation between aspects quite sufficiently, meaning that it offers "no account of the individual as explanatory whole" (Kyselo, 2014, p. 1). In other words, despite acknowledging the existence of meaningful dynamical relations between self-patterns, Gallagher's account "doesn't develop a full theory about how the various elements of the pattern of self are connected" (Beni, 2016, p. 3,731). Although these objections are directed at the pattern theory in the first place, they also bear on the issue of the extendedness of the self. To support an extended conception of the self we need to accept that the self always latches onto the ecological and/or social environment, and the unit of analysis is the self-environment (Gallagher, 2013, p. 4). But if the pattern theory fails to produce a full account of the relationship between different aspects of the self, trivially it would fail to explain how extended (and situated) contributors are indeed component parts of the self-pattern in a constitutive sense. Accordingly, the pattern theory would fail to account for the extendedness of the self-pattern. This could be a significant blow to the extended cognition thesis as well as a general objection to the pattern theory. That said, I have to immediately add that Gallagher and Daly offer some interesting strategies for accounting for the relationship between diverse aspects of the self. The most promising of their suggested strategies (in my opinion) consists of invoking predictive processing and the free energy principle. Why is this the case? Patterns that are at issue in the pattern theory are specified in terms of dynamical system theory. Gallagher's insight into that subject receives support from some important works such as (Schöner and Kelso, 1988; Kelso, 2016). However, this paper assumes that it could be *also* worthwhile to invoke comprehensive and unifying formal framework under which to model relations between diverse aspects of the self (as well as the relation between the self and the environment). The Free Energy Principle (FEP) seems to underpin such a comprehensive, unifying framework. The dynamic approach too endeavors to account for the emergence of the patterned behavior under generative self-organizing processes (Schöner and Kelso, 1988; Kelso, 2016). FEP can be used in the same spirit to achieve the same goal with remarkable formal precision and empirical success. In view of the mathematical vigor and empirical success of the FEP, it seems that FEP provides a suitable theoretical framework for bolstering Gallagher's account of the relationship between aspects of self-patterns.

The Free Energy Principle (FEP) and predictive processing, characterized in terms of Bayesian models of minimization of variational free energy, are the unifying theoretical framework that accounts for perception, cognition, and action (Friston, 2010; Hohwy, 2013; Clark, 2016). In order to survive, organisms must remain in *non-equilibrium steady states*. This means that they must avoid getting into unpredicted situations. The probabilistic description of the dynamics of systems in non-equilibrium steady states is developed into two kinds of descriptions. According to Ramstead et al. (2020, p. 6):

> First, the system can be described in terms of the flow of the system's states—that are subject to random fluctuations—in which case, we can formulate the flow in terms of a path integral formulation, as a path of least action. Equivalently, we can describe the non-equilibrium steady-state in terms of the probability of finding the system in some state when sampling at any random time.

According to this formulation, self-organizing systems (in terms of intrinsic geometry) evolve toward some non-equilibrium steady-state density which can be interpreted as a statistical or generative model (in terms of its extrinsic geometry). In this fashion, we could characterize the joint probability density over internal states and external states (Ramstead et al., 2020, p. 9). Within this context, variational free energy is an information-theoretic measure that provides an upper bound on surprise. Entropy is "[t]he average surprise of outcomes sampled from a probability distribution or density" (Friston, 2010, p. 1). Living systems minimize their free energy by staying in a small set of environmental states. A fish needs to stay in the water because a fish out of water will find itself in a surprising state. Staying in a limited number of states enables organisms to form approximately precise predictions of the environment. This makes the organisms' interactions with the environment efficient. The organism can minimize its free energy either via adjusting its models (that's predictive coding) or via action (that's active inference). It can minimize its free energy by either changing its internal models of the environment based on evidence that is sampled actively or by acting on the environment and changing the environmental states to make them match its predictions. When applied to the brain, the theory holds that the brain could get approximate representations of the causal structure of the environment by minimizing prediction errors[3]. Below, I shall unpack this remark.

The brain forms generative models[4] of the environment and through top-down processing in a hierarchical organization represents the real world. In case of discrepancy between predictions and actual sensory inputs, the brain minimizes its prediction errors and finesses its generative models (or the organism changes the environmental states to match the predictions) (Friston and Stephan, 2007). FEP and predictive processing are used to provide viable models of selfhood (Limanowski and Blankenburg, 2013; Apps and Tsakiris, 2014; Limanowski and Friston, 2020). At least for some organisms, having a representation of the self in generative models is indispensable to the multisensory integration in

[3]Not all representatives of predictive processing would agree to using "representations" in this context. A radical embodied approach would deny that internal models or inner simulacrums play a significant role in PP. But moderate advocates of embodiement such as Clark conceds that models (which embed representations) do not need to be totally eliminated from predictive processing. According to Clark's moderate version of embodiment, "it is surely that very model-invoking schema that allows us to understand how it is that these looping dynamical regimes arise and enable such spectacular results" (Clark, 2016, p. 293).

[4]Generative models are internal probabilistic models that the brain uses to update its posterior models.

both exteroceptive and interoceptive streams. On such grounds, Gallagher and Daly's (2018, p. 8) argue that FEP and predictive processing characterize the dynamical relations that bring together otherwise diverse self-patterns. Let us see how this affects the extendedness of the self.

Because there are dynamical relations between self-patterns, it can be assumed that the extended aspect is somewhat connected to other aspects of the self. But does this mean that the extended aspect is a *constituent* of the self (in contrast, it could be assumed that it is related to other self-aspects loosely and without forging any strong ontological bonds? Gallagher and Daly's elaboration on dynamical relations between self-patterns is silent about this. Moreover, aside from a fleeting reference to Dennett's (1991) theory of real patterns, Gallagher and Daly do not explicate their view on the existence of the self-pattern. The question of (modes of) the existence and reality of the self needs to be treated with adequate technical tools.

Gallagher and Daly's characterization of dynamical relations between aspects of the self indicates that Gallagher is not committed to the existence of a class of totally diversified and disintegrated self-contributors. Nor does he conceive of the self-pattern in terms of a classical substance. This puts the ontological status of the self-pattern in a twilight zoon. Inspired by Gallagher and Daly's, (2018 p. 2) remark on the Dennettian tendency of their view, I suggest that the self is a scattered composite pattern that is constituted by diverse aspects, the extended aspect included. I use metaphysical tools that are congenial to Dennett's (1991) theory of *real patterns* to substantiate my stance on the existence and reality of the self as a composite pattern. It is true that at times Dennett seems something of a pragmatist about the reality of the pattern, and doesn't offer any heavy ontology[5]. However, Dennett (1983, p. 380) is clear that he is not a fictionalist about theoretical posits such as the center of gravity. This is because these posits play an explanatory function (and thus could be embraced based on some indispensability argument). The result is a moderate metaphysical stance that cannot be described in simple terms of realism vs. instrumentalism. According to Dennett, his real patterns theory "is clearer than either of the labels [meaning realism and instrumentalism]," so he just leaves "that question to anyone who still finds illumination in them" (Dennett, 1991, p. 51). This approach is in harmony with Petersen's take on algorithmic metaphysics. Petersen submits that "I must confess that I am sympathetic not only to Dennett's patternist proposal, but also to this metaontological stance [of Dennett's, which has been just cited]" (Petersen, 2019, p. 3). I suggest that this metaphysical enterprise can be applied to deal with the question of the extendedness of the self-pattern.

More light will be shed on this topic if we ponder the two following questions:

1. Is the extended aspect constitutive of the self-pattern?
2. Do the self-pattern and the environment constitute a genuine composite object?

We need to know more about the metaphysics of composed patterns before providing viable answers to these questions.

[5]I thank one of the reviewers of this journal for reminding me of this point.

AN ALGORITHMIC METAPHYSICS OF COMPOSITION

We can address the question of how to draw the boundaries of a cognitive system if we could tell when two systems that are coupled form an integrated system. This question resembles the question of *composition*, which asks when we can claim that some objects constitute a new object. This paper takes a compositional stance on constitution.

Generally, the question of the composition provides metaphysical insights into the thesis of extendedness. It may be assumed that there are no composite objects at all, or it may be assumed that any mereological sum constitutes an integrated object. Between these two extremes, there are moderate varieties; some pluralities (such as atoms of hydrogen and oxygen) constitute a new object (such as a molecule of water) and some other pluralities (such as the compound of the pear tree in my yard and the Taj Mahal) do not constitute a new object[6]. In this context, Petersen is advocating a compositional conception of constitution (Petersen, 2013, p. 312). According to this approach, for an object to be *constituted*/*composed* by some pluralities, there must exist some degree of "connectedness" or "integrity" between the pluralities (Simons, 2000, p. 290). Integrity and connectedness are conditions that need to be satisfied by constitution. This is because without integrity and connectedness the aggregates would be assembled into an arbitrary sum. In this sense, I adopt a compositional stance on constitution (Mark the similarity of the problem of composition/constitution to the problem of the relation between self-contributors and aspects. The general insight of this paper is that from a metaphysical point of view, the self can be identified with a scattered composite pattern).

There have been significant attempts at invoking information-theoretic frameworks for identifying the structure of reality, or more technically, real patterns (Dennett, 1991; Ross, 2000; Ladyman and Ross, 2007). According to Dennett's statement of the patternist approach, "A pattern exists in some data-is-real-if there is a description of the data that is more efficient than the bit map, whether or not anyone can concoct it" (Dennett, 1991, p. 34). Interestingly enough, Dennett's conception of real patterns is in line with Gallagher and Daly's conception of self-patterns [Referring to Dennett's pattern theory, Gallagher indicates that "the self has the scientifically useful reality of a pattern" (Gallagher and Daly's, 2018, p. 2)]. I shall flesh out this

[6]To the question composition, van Inwagen provides a simple answer in terms of organicism, which holds that "the activity of the xs constitutes a life or the xs are the current objects of a history of maintenance" (van Inwagen, 1995, p. 138). Xs that constitute a life do compose exactly an organism (ibid, p. 91). Of course, the organicist criterion of composition can lead to a rough and ready answer to the question of how to draw the boundaries of cognitive systems—obviously by laying the boundaries of cognition (or composition of the object) on the boundaries of the organism's body. But for one thing, the criterion precludes the possibility of extended cognition into non-organismic objects too trivially (perhaps based on an unsubstantiated prejudice in favor of being organic). Moreover, organicism may be construed to indicate that only living organisms and mereological simples exist, but there are no non-living composite objects such as tables and chairs. This view, called "the denial" by van Inwagen (1995, p. 1) is too radical to be justified easily.

proposal with an eye to its use for dealing with the question of the self (as a composite reality) and its metaphysical aspects. This proposal draws a connection between the metaphysical definition of objects (as patterned or structured entities) and their description in terms of compressibility (compressible objects are patterned). Patterns that are described by compressed programs are indispensable to viable representations of the world. Dennett generally implies that the patterns could be characterized based on Kolmogorov complexity. James Ladyman and Don Ross have used the ideas of logical depth[7] and projectibility to characterize the patterns. Projectible patterns, say Ladyman and Ross, are real patterns in the dataset[8,9]. Petersen (2013, 2019) characterizes real patterns by invoking algorithmic information theory (and more specifically, in terms of Kolmogorov complexity).

Petersen's goal is to show how the patternist approach can make sense of "composition" which is metaphysically a vague and mysterious notion. Kolmogorov complexity [K (x)] efficiently represents the main insight behind Dennett's and Ladyman and Ross's views on the representation of reality. Here, the notion of incompressibility (in terms of Kolmogorov complexity or logical depth, which are arguably translatable to one another) provides a criterion of the constitution of an object, given that "To be is to be a real pattern" (Ladyman and Ross, 2007, p. 233). The relevance of the present discussion to the issue of extendedness is this: the criterion of being a real composite could be used to determine whether the biological cognitive system and the environment form a real composite entity. More specifically, I argue that the criterion can be used to make sense of the integrity and connectedness of the extended aspect with the rest of the self-pattern. We need to show that the extended aspect is constitutive of the self. Then we can conclude that the self-pattern is minimally extended. Below, I shall furnish more details about the criterion of being a composite patterned object.

According to Petersen (2013, 2019) the criterion of being a real pattern (characterized in terms of Kolmogorov complexity) can demarcate what is a genuine composite object from the mere sum of independent objects or patterns. According to Petersen's proposal, an aggregate of objects is itself a real object if there is some kind of integrity and connectedness between its component parts. In other words, real composite objects are simpler than the sum of their independent component parts. In this fashion, Kolmogorov complexity can be incorporated into an ontological criterion of what is real. According to Petersen, given that "compressibility" corresponds to "simplicity," there is ontological gain when there is some gain in a pattern. This definition provides insights into the internal integrity of genuine composite objects. This is because "to compose, a compressible region must be referenced by the best compression of the totality in which the region resides" (Petersen, 2019, p. 10). I unfold the technical details immediately.

Complexity and simplicity are defined in terms of the processing of information in a universal Turing machine, which is an abstract device that can model any computable algorithm in a discrete domain. A Turing machine is constituted by a finite program. It can manipulate a tape (which is a linear list of cells), and it has a head. The machine can fill each cell with any of the symbols from a specified set of variables, and it can move the head to any specific cell. Based on such simple operations, a Turing machine can model everything in the discrete domain that is intuitively computable. A universal Turing machine can model the behavior of any other Turing machine (Vitanyi, 2009). The relation between the notions of "Turing computation" and "Kolmogorov complexity" is this: Kolmogorov complexity of an object consists of the length of the shortest program (i.e., shortest input) that produces that object, assuming that the program is processed by a fixed universal prefix Turing machine [not all theorists agree that the domain of computable should be discrete (Hutter, 2008)]. The Turing machine program is the description of that object, and an object that has such a shortest description is considered to be simple. Technically, for the string x, the program p provides the shortest description, if when processed by the universal Turing machine U P outputs x. Under that supposition, the shortest description is provided by

$$K_U(x) : = min_p\{l(p) : U(p) = x\}$$

l(p) submits the length of p in bits (Hutter, 2008). The definition of complexity that is at issue here is compatible with the definition of logical depth as stated above. And Petersen builds upon the formal definition of Kolmogorov complexity to address the metaphysical question of objecthood in a world that includes some fundamental objects and simple properties (this world also accommodates the succession of time). The question is this: could there be composite objects in this world. This leads us to another important question: what is the criterion of demarcating genuine composite objects from compounds that do not constitute genuine objects. To find answers to these questions, Petersen develops an algorithmic-compositional concept of "constitution" for both objects and their properties (Petersen, 2013, p. 312). I cite Petersen (2013, pp. 308–309) to show how Kolmogorov complexity is developed into a criterion of being a composite object. Let $l \in L$ be any interval and x_l be a composite function

[7]Logical depth is defined as "a normalized quantitative index of the execution time required to generate the model of the real pattern in question by a near incompressible universal computer program, that is, one not itself computable as the output of a significantly more concise program" (Ladyman and Ross, 2007, p. 220).

[8]According to this proposal:
To be is to be a real pattern; and a pattern x → y is real iff

> (i) it is projectible; and
>
> (ii) it has a model that carries information about at least one pattern P in an encoding that has a logical depth less than the bitmap encoding of P, and where P is not projectible by a physically possible device computing information about another real pattern of lower logical depth than x → y (Ladyman and Ross, 2007, p. 233).

[9]It has been contended that this criterion of projectibility cannot demarcate *real* patterns (or at least partial non-redundant patterns) from patterns simpliciter (Beni, 2017; Suñé and Martínez, 2019). But these considerations do not deter us from continuing our pursuit, because our present enquiry is not concerned with the association between non-compressibility and *reality* (more on this later in the paper).

restricted to that interval. $x^{\#}{}_l$ designates the length of x plus some small constant that denotes the computational overhead;

> *x_l is a **composite object** if and only if*
> *1. $KU(x_l) < x^{\#}{}_l$ (the **compressible clause**).*
> *2. There is no partition of l into intervals $\{l_1 \ldots l_n\}$ such that $\sum_i KU(xli) \leq KU(xl)$ (the **minimal clause**).*
> *3. There is no interval l' containing l such that $KU(xl') \leq KU(xl)$ (the **maximal clause**).*

Minimal clause indicates that If x has a sub-region that does not contribute to the compressibility of the remainder, then the diverse components in x do not constitute anything (Petersen, 2019, p. 13). Consider two objects (such as the pear tree and the Taj Mahal, call them o_1 and o_2) that do not constitute a genuine composite object (call it o_3 which is equal to $o_1 \cup o_2$). As Petersen argues, although o_3 is compressible, it is not an object because of the minimal clause, given that $KU(o_1)+KU(o_2) \leq KU(o_3)$ (Petersen, 2013, pp. 309–310). Thus, the union of the pear tree and the Taj Mahal is not a genuine object. On the other hand, maximal clause indicates that "parts must each be simpler than wholes, but wholes must be simpler than all their parts taken separately" (Petersen, 2019, p. 15). Consider the possibility of breaking the program that describes o_1 into two substrings o_{1R} and o_{1L} (representing the right and left substrings). That is to say, $o_1 = o_{1R} \cup o_{1L}$. It might be assumed that the same kind of argument that was mentioned to rule out o_3 as a genuine object could be used to rule out o_1 too, by indicating that it runs afoul of the minimal clause (instead o_{1R} and o_{1L} are genuine objects). The branches and the trunk of the same pear tree (or its atoms) could be modeled as separate objects, and it could be assumed that the pear tree itself is not a genuine object. The maximal clause excludes this option by preventing the arbitrary division of proper objects. That is to say, o_{1R} and o_{1L} are not proper objects. Although in principle we may be able to decompose the pear tree into the independent classes of its branches and its trunk, there is no gain in simplicity or ontology of decomposing the tree in this fashion. Let us see how this applies to the question of constitution of the self and its extendedness.

IS THE SELF A COMPOSITE OBJECT?

Petersen (2019, p. 5) submits that "Seeking to minimize Bayesian surprise on higher-order parameters is basically just pattern extraction." FEP is stated in terms of Shannon information theory (rather than Kolmogorov complexity). But there are formal links between Shannon information theory and Kolmogorov complexity (Grunwald and Vitanyi, 2004). At any rate, FEP and predictive processing are used to characterize the self. The self is a composite pattern. It is compressible in sense of Kolmogorov complexity and logical depth.

According to Gallagher, the self does not have an independent existence. But from the point of view of algorithmic metaphysics, the question is not about *the independent existence* of the self but the composition of the self. This is in line with a compositional metaphysical view. The question that we must attend to is this: *Is the self-pattern simpler or more compressible than the sum of its independent contributors*. A positive answer to this question indicates that aspects are constituting the self, instead of loosely hanging together, and it would follow that the self has a genuinely extended aspect. This follows from the application of the metaphysical criterion of compositionality. Below, I explain how these three clauses apply to the issue of extendedness of the self.

As to the first clause, it could be easily granted that the self-pattern is compressible. But does this mean that the self is a composite pattern? According to the minimal clause, if the self is a genuine composite object, the sum of independent aspects of the self cannot be simpler (or more compressible) than the self. It could be the case that the extended aspect and the cognitive aspect are each simpler than the self as a composite object, but the sum of all involved self-contributors is not simpler or less complex than the self-pattern (see the maximal clause in the previous section). Together, the minimal and maximal clauses indicate that for the self to be a genuine composite object (or pattern), its description must be simpler and shorter than the sum of descriptions of diverse self-aspects and contributors. There is nothing in the definition of self-patterns that preclude this possibility. Take Newen et al.'s conception of patterns, which is adopted by Gallagher (2013):

> A feature F is constitutive for a pattern X if it is part of at least one set of features which is minimally sufficient for a token to belong to a type X. "Minimally sufficient" means that these features are jointly sufficient for the episode to be of type X, but if one of them were taken away the episode would no longer count as an instance of X (Newen et al., 2015, p. 195).

This definition does not indicate that separate aspects have their independent existent, or the sum of independent self-contributors is more endurable, compressible, or simpler than the self as a composite pattern. That is to say, although the self does not have an independent existence, self-aspects are even less capable of having their independent existence. In this sense, it could be assumed that self-aspects are constituting the self, and the self-pattern is ontologically more fundamental than separate self-aspects. This provides an insight into the constitution of the self. A more technical demonstration can be offered in terms of the FEP-based characterization of the self.

FEP is indeed formulated in terms of Shannon information theory, rather than Kolmogorov complexity[10]. Even so, the FEP-based account considers the self as a theoretical posit that can reduce the complexity of our explanation of various aspects and elements. The sum of separate self-aspects cannot explain cognition and action of a person in a simple and unified way. This means that the sum of explanations that diverse self-aspects produce is more complex than their integrated explanation under the rubric of FEP. The general insight

[10]The difference between Shannon theory of information (which provides the theoretical foundation of FEP) and Kolmogorov complexity is that the former models the randomness of the source of information whereas the latter describes the randomness of the object itself (Grunwald and Vitanyi, 2004, p. 3).

here is that the self is formed around the idea that "one's own body is the one which has the highest probability of being 'me' as other objects are probabilistically less likely to evoke the same sensory inputs" (Apps and Tsakiris, 2014, p. 6). Therefore, stipulating the self as a theoretical posit maximizes the simplicity of cognitive and biological mechanisms by minimizing the overall information conveyed in the system (that is the entropy[11] of the system that represent the distribution of probabilities that represent the structure of the environment). In words of Apps and Tsakiris, "the notion that there is a 'self' is the most parsimonious and accurate explanation for sensory inputs. In mathematical terms, this parsimonious accuracy is exactly the quantity that is optimized when minimizing free energy or prediction error" (Apps and Tsakiris, 2014, p. 89). Gallagher and Daly build upon this fundamental insight to substantiate their view on the existence of meaningful dynamical relations between diverse self-aspects[12].

A high-level description of the self as a unified entity can explain how minimizing the discrepancy between the generative models and the environment (and one's own body) generates perception and cognition. The sum of independent self-aspects fails to explain the organism's representational and active capacities with the same amount of simplicity and fruitfulness. If that is true, then the self is more than just a cluster concept (as Gallagher's original pattern theory in 2013 paper indicates). Self indeed lacks an independent existence, but it contributes to simpler explanatory schemes in ways that remain beyond the sum of diverse self-aspects.

Finally, it is worth mentioning that Petersen's definition of composite objects allows for the existence of scattered objects. For example, it indicates that although there are no strong bonds between water molecules that constitute a cloud, the cloud can be recognized as a composite object, albeit a scattered one (Petersen, 2013, p. 311, 2019, p. 8). In this fashion, the self can be identified as a scattered composite pattern. This is because stipulating the self leads to a less complex description of the organization of multiple self-aspects. I shall unfold the consequences for the extendedness of the self.

From our discussion in this section, it follows that the extended aspect is not just loosely hanged to the aggregate of other independent self-aspects. The extended aspect, along with other elements, constitute the self. This means that the self is *minimally* extended. The extended aspect is not just *coupled* with other aspects, they *constitute* a genuinely composite entity, in the sense that is at issue in the compositional view on constitution.

It is worth repeating that Gallagher takes an enactivist stance on the question of constitution, and explicates it in terms of "reciprocal causal relations" (Gallagher, 2018). Accordingly, Gallagher ignores the coupling-constitution fallacy and takes the viability of the extended mind approach for granted. While I do not challenge the validity of the enactivist stance, I do not think philosophical fundamental stances would be justified, confirmed, or verified easily. One can embrace them by pondering a number of various considerations, such as simplicity, fruitfulness, etc. While I do not challenge the general plausibility of the enactivist stance, but I think the compositional view deserves to be taken seriously too. Gallagher's theory does indicate that the self includes an extended embodied aspect, albeit without appealing to a compositional criterion of constitution. It might indeed be possible to understand the cluster concept of the self (which also embeds an extended aspect) in terms of a dynamical gestalt, constituted by reciprocal causal relations (and thereby by a coupling relation with the environment) rather than compositionality)[13]. This paper does not aim to refute the enactivist approach. It only aspires to provide a metaphysically well-posed alternative to it. This is stated in terms of a criterion of compositionality, and it has the edge over the dynamical systems approach in the following way: the dynamical system approach cannot set a meaningful distinction between causally related clusters that do constitute an object (such as the self) from causal-coupling relations that do not constitute an object (such as the compound of the self and the environment). The compositional approach can set such a distinction. I understand that the advantage that I attribute to the compositional approach may not persuade the enactivist to embrace my proposal. I simply state the compositional criterion to argue that the self and the environment constitute a composite system, without claiming the absolute superiority of this construal over enactivism (the paper is rather unassuming in this sense). In the next section, I will consider the question of the extendedness of the self by asking whether the "self-environment" is a genuine composite entity.

DO THE SELF AND THE ENVIRONMENT CONSTITUTE AN OBJECT?

As we have already seen, the self can be characterized in terms of FEP and predictive processing. There are ecological and enactivist construals of predictive processing and active inference (Bruineberg et al., 2016; Gallagher and Allen, 2016). Predictive processing and FEP are concerned with the state of homeostasis, which is the state of stable internal equilibrium of the organism with the environment. The ecological construal of FEP and predictive processing represent the relation between the organism and its environment in terms of dynamical

[11] Formally, entropy is defined in terms of the amount of information that an observer would gain after receiving a given message. For a random variable X, Shannon entropy is defined as:

$$HX = \sum_{x \in X} px \log 1/px$$

[12] Once more, please note that because FEP and predictive processing are stated in terms of Shannon information theory, they are not concerned with the complexity of the object (so much as the source of information). However, FEP conveys clear implications about the simplicity that the assumption of the existence of the self brings to the explanation of cognition and action (not the same could be told of diverse self-patterns).

[13] I thank one of the reviewers of this journal for pointing out this to me.

coupling of the organism with the eco-niches and its windows of affordance. There are also ecological and enactivist theories of the self and "mineness" in terms of active inference under FEP (Kiverstein, 2018). The question of extendedness of the self is this: Do the self and the environment constitute a genuine composite object, or they are just coupled together? The point that "[t]he organism embodies in its biological organization a hierarchically structured model of its *own* existence in its environment, or equivalently its being-in-the-world" (Kiverstein, 2018, p. 2) could be appreciated rather easily in the context of the pattern theory of the self. This is because (according to what we saw in the previous section) the self-pattern includes an extended aspect. But this is not quite enough for establishing the point that the self and the environment are component of a genuine composite object. I shall clarify immediately.

Self-organizing systems are minimizing their free energy by garnering evidence for their inbuilt generative models. They are self-evidencing in the sense that they endeavor to actively garner evidence for their existence (Hohwy, 2014). And some of these self-evidencing organisms are specified as "selves" or as "subjects of minimal phenomenal experiences" under FEP (Limanowski and Blankenburg, 2013; Kiverstein, 2018). One way of fleshing out this is by assuming that FEP can be used to describe the self as a subject of phenomenal experience. Selves are capable of modeling their *expectations* about the *future* states and the *consequences* of their actions (Friston, 2018, p. 579). Selves (as subjects) can model different consequences of their actions for themselves and choose one particular course of action amongst several possible ones (Friston, 2018, p. 6). In other words, we need to have models of ourselves as trajectories with non-linear effects on our sensory input. This accounts for perceptual unity in a wide time-perspective (Hohwy, 2013, chapter 10). According to Hohwy:

> Action arises when prediction error minimization happens by acting on the world while sticking with one's counterfactual about the world. For this kind of strategy to be feasible we need an ordering of policies for how to go about minimizing error in this way. Such policies are expectations about how flows of error are minimized as we move through the world. These expectations must rely on hypotheses under a hierarchical model of ourselves including our own mental states as coherent and unitary causal trajectories (Hohwy, 2013, p. 255).

In this vein, a self-conscious system is defined as "a system that can simulate multiple futures, under different actions, and select the action that has the least surprising outcome" (Friston, 2018, p. 5). But does this mean that the self is a component of a genuine composite entity (call it the self-environment compound), in a way that is demanded by a strong version of the extended thesis?

A strong version of the extended thesis can be stated like this: the self is extended to the environment, and the self-environment compound is constituted by both the environment and the self as its constituents. If so, the existence of the self depends on its role as a constituent of the self-environment compound. To substantiate this claim within the patternist framework we must be able to show that the self-environment compound is simpler or more compressible than the sum of the self and the environment as independent entities.

Let us grant that the self-environment compound is compressible (this means that we can grant the compressible clause). However, it is not the case that the self-environment compound is simpler or has a more independent existence than the sum of independent components—namely the self and the environment. I shall unfold this remark immediately.

The self and its aspects are described via Markovian models (Friston et al., 2020; Parr et al., 2020). Markov blankets are networks in Bayesian spaces that register a separation between sensory states and active ones, given that sensory states are independent of internal states and active states are independent of external states. Friston et al. (2020) weaved Markovian blankets into a framework of information geometry. The result is an informational/probabilistic description of the way that the brain represents the external world to itself based on the relationship between "probability distribution *about* things" and "probability distribution *of* things" (Friston et al., 2020). Their description of the brain-world relationship accommodates representation of *expected surprise* as the set of beliefs that organisms hold about the consequences of their actions in the world (this provides a basis for phenomenal aspects of the minimal self). Not only Markovian models (with separable internal and external spaces) describe the relationship between the brain and the world, they also model notions of agency, consciousness, and deliberate pre-meditated action (as the properties of the minimal self).

To return to the discussion of compressibility (and minimal and maximal clauses), when constructing their models of consciousness and agency Friston et al. (2020), employed an information geometry that includes a metric for measuring the distance to informational states space. Why this is relevant to the issue of compressibility? Because the information-theoretic measure provides a formal criterion for dealing with the question of compressibility and simplicity. To explicate compressibility and simplicity in information geometry, we should consider the following question. Which of the two classes of entities is simpler or more compressible? The self-environment compound or the sum of the self and the environment as separate entities? Not only the self-environment as a composite entity is less simple than the self and the environment (and their sum), the formal statement of the self and its phenomenal aspects indicates that they are not constituents of the self-environment compound (in the compositional sense). Using Markovian models indicates that to be modeled, the self, as a self-evidencing organism, must be described as an entity with rather clear boundaries that separate it from its environment.

On the same subject, an advocate of the enactivist, extended mind approach does not need to assume that Markov blankets are only in the business of separating inside from outside. Markov blankets can be also used to show how inside and

outside are connected (or coupled). Once more, the disagreement about the issue of "constitution" and its relation with "causal coupling" raises its head. For the enactivist, who assumes that causal coupling is enough for the constitution, Markovian models are venues of extension of cognition (and selfhood). However, for those who advocate the compositional view of constitution, the coupling relation is not enough for establishing the extendedness of the self.

The significance of the barrier or the evidentiary boundary between the self and the environment has been emphasized by Hohwy (2007, 2013). The point about the use of Markovian models in describing the brain-world relationship cements the importance of the barrier between the self with its environment (Hohwy, 2017; Kirchhoff et al., 2018). It is possible to see Markov blankets as the venue of dynamical interaction between the organism and its environment. Even so, there is a solid construal which represents Markov blankets as separating boundaries that seclude the organism (or its *self*) from the environment. Although the self is not completely secluded from the environment by boundaries of skin and skull (Kirchhoff et al., 2018; Kiverstein and Rietveld, 2018), the use of the Markov blanket implies that there are staunch boundaries between the self and its environment. This is in line with Hohwy's (2013, 2014, 2017) representationalist construal of FEP. According to this construal, the brain is secluded from the world, and it infers the state of the world from beyond an inferential veil. The Markov blanket here sets robust boundaries that separate the self from its environment. Aside from Hohwy's construal, Friston and colleagues have suggested that the Markov blanket could contribute to separating boundaries (Friston et al., 2020; Parr et al., 2020). This latter construal (which presents the Markov blanket as a dividing boundary) is in harmony with assuming that the self and the environment do not constitute a non-decomposable composite entity.

CONCLUDING REMARKS

The paper invoked the criteria of simplicity and complexity of real patterns to deal with two specific questions.

1. Is the extended aspect constitutive of the self-pattern? If this is the case, there is some purchase for the extendedness of the self under the pattern theory.
2. Do the self and the environment constitute a genuine composite object?

Applying the criteria that are drawn from Petersen's algorithmic metaphysics, I argued that while there is some basic purchase for arguing that the self is minimally extended, the self and the environment do not constitute a real composite object.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Adams, F., and Aizawa, K. (2008). *The Bounds of Cognition: The Bounds of Cognition*. Hoboken, NJ: Blackwell Publishers, doi: 10.1002/9781444391718

Adams, F., and Aizawa, K. (2009). "Why the mind is still in the head," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (Cambridge: Cambridge University Press), 78–95. doi: 10.1017/CBO9780511816826.005

Adams, F., and Aizawa, K. (2012). *Why the Mind Is Still in the Head: The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press, 78–95. doi: 10.1017/cbo9780511816826.005

Aizawa, K. (2010). The coupling-constitution fallacy revisited. *Cogn. Syst. Res.* 11, 332–342. doi: 10.1016/j.cogsys.2010.07.001

Apps, M. A. J., and Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* 41, 85–97. doi: 10.1016/J.NEUBIOREV.2013.01.029

Barsalou, L. W. (2008). Grounded cognition. *Ann. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Beni, M. D. (2016). Structural realist account of the self. *Synthese* 193, 3727–3740. doi: 10.1007/s11229-016-1098-9

Beni, M. D. (2017). Structural realism, metaphysical unification, and the ontology and epistemology of patterns. *Int. Stud. Philos. Sci.* 31, 285–300. doi: 10.1080/02698595.2018.1463691

Beni, M. D. (2019a). *Structuring the Self*. London: Palgrave Macmillan.

Beni, M. D. (2019b). An outline of a unified theory of the relational self: grounding the self in the manifold of interpersonal relations. *Phenomenol. Cogn. Sci.* 18, 473–491. doi: 10.1007/s11097-018-9587-6

Bitbol, M., and Gallagher, S. (2018). The free energy principle and autopoiesis. *Phys. Life Rev.* 24, 24–26. doi: 10.1016/j.plrev.2017.12.011

Bruineberg, J., Kiverstein, J., and Rietveld, E. (2016). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese* 16, 1–28. doi: 10.1007/s11229-016-1239-1

Carter, J. A., and Kallestrup, J. (2019). Varieties of cognitive integration. *Noûs* 19:nous.12288. doi: 10.1111/nous.12288

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Clark, A. (2016). *Surfing Uncertainty*. Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780190217013.001.0001

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 7–19. doi: 10.1093/analys/58.1.7

Dennett, D. C. (1983). Intentional systems in cognitive ethology: the "Panglossian paradigm" defended. *Behav. Brain Sci.* 6, 343–355. doi: 10.1017/S0140525X00016393

Dennett, D. C. (1991). Real patterns. *J. Philos.* 88, 27–51.

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nat. Rev Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K. J. (2018). Am I self-conscious? *Front. Psychol.* 9:579. doi: 10.3389/FPSYG.2018.00579

Friston, K. J., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y

Friston, K. J., Wiese, W., and Hobson, J. A. (2020). Sentience and the origins of consciousness: from cartesian duality to markovian monism. *Entropy* 22:516. doi: 10.3390/E22050516

Gallagher, S. (2013). A pattern theory of self. *Front. Hum. Neurosci.* 7:443. doi: 10.3389/fnhum.2013.00443

Gallagher, S. (2018). "New mechanisms and the enactivist concept of constitution," in *The Metaphysics of Consciousness (207–220)*, ed. M. P. Guta (London: Routlege), 207–220. doi: 10.4324/9781315104706-13

Gallagher, S., and Allen, M. (2016). Active inference, enactivism and the hermeneutics of social cognition. *Synthese* 16, 1–22. doi: 10.1007/s11229-016-1269-8

Gallagher, S., and Daly, A. (2018). Dynamical relations in the self-pattern. *Front. Psychol.* 9:664. doi: 10.3389/fpsyg.2018.00664

Grunwald, P., and Vitanyi, P. (2004). Shannon information and kolmogorov complexity. *arXiv* [Preprint]. http://arxiv.org/abs/cs/0410002 (accessed October 10, 2019).

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche* 13, 1–20.

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780199682737.001.0001

Hohwy, J. (2014). The self-evidencing brain. *Noûs* 50, 259–285. doi: 10.1111/nous.12062

Hohwy, J. (2017). "How to entrain your evil demon," in *Philosophy and Predictive Processing*, eds T. Metzinger and W. Wiese (Frankfurt: MIND Group), doi: 10.15502/9783958573048

Hutter, M. (2008). Algorithmic complexity. *Scholarpedia* 3:2573. doi: 10.4249/scholarpedia.2573

Kelso, J. A. S. (2016). On the self-organizing origins of agency. *Trends Cogn. Sci.* 20, 490–499. doi: 10.1016/j.tics.2016.04.004

Kirchhoff, M. D. (2015). Extended cognition & the causal-constitutive fallacy: in search for a diachronic and dynamical conception of constitution. *Philos. Phenomenol. Res.* 90, 320–360. doi: 10.1111/phpr.12039

Kirchhoff, M. D., and Kiverstein, J. (2019). *Extended Consciousness and Predictive Processing: A Third-Wave View*. Abingdon: Routledge.

Kirchhoff, M. D., Parr, T., Palacios, E., Friston, K. J., and Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *J. R. Soc. Interface* 15:20170792. doi: 10.1098/rsif.2017.0792

Kiverstein, J. (2018). Free Energy and the self: an ecological–enactive interpretation. *Topoi* 39, 559–574. doi: 10.1007/s11245-018-9561-5

Kiverstein, J., and Rietveld, E. (2018). Reconceiving representation-hungry cognition: an ecological-enactive proposal. *Adapt. Behav.* 26, 147–163. doi: 10.1177/1059712318772778

Kyselo, M. (2014). The body social: an enactive approach to the self. *Front. Psychol.* 5:986. doi: 10.3389/fpsyg.2014.00986

Ladyman, J., and Ross, D. (2007). *Every Thing Must Go*. Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780199276196.001.0001

Limanowski, J., and Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Front. Hum. Neurosci.* 7:547. doi: 10.3389/fnhum.2013.00547

Limanowski, J., and Friston, K. (2020). Attenuating oneself. *Philos. Mind Sci.* 1:6. doi: 10.33735/phimisci.2020.I.35

Metzinger, T. (2003). *Being no One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

Newen, A. (2018). The embodied self, the pattern theory of self, and the predictive mind. *Front. Psychol.* 9:2270. doi: 10.3389/fpsyg.2018.02270

Newen, A., Welpinghus, A., and Juckel, G. (2015). Emotion recognition as pattern recognition: the relevance of perception. *Mind Lang.* 30, 187–208. doi: 10.1111/mila.12077

Parr, T., Da Costa, L., and Friston, K. J. (2020). Markov blankets, information geometry and stochastic thermodynamics. *Philos. Transact. R. Soc. A* 378:20190159. doi: 10.1098/rsta.2019.0159

Petersen, S. (2013). *Toward an Algorithmic Metaphysics*. Berlin: Springer, 306–317. doi: 10.1007/978-3-642-44958-1_24

Petersen, S. (2019). Composition as pattern. *Philos. Stud.* 176, 1119–1139. doi: 10.1007/s11098-018-1050-6

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., and Spivey, M. J. (2012). Computational grounded cognition: a new alliance between grounded cognition and computational modeling. *Front. Psychol.* 3:612. doi: 10.3389/fpsyg.2012.00612

Piredda, G. (2017). The mark of the cognitive and the coupling-constitution fallacy: a defense of the extended mind hypothesis. *Front. Psychol.* 8:2061. doi: 10.3389/fpsyg.2017.02061

Ramstead, M. J. D., Friston, K. J., and Hipólito, I. (2020). Is the free-energy principle a formal theory of semantics? From variational density dynamics to neural and phenotypic representations. *Entropy* 22:889. doi: 10.3390/e22080889

Ross, D. (2000). "Rainforest realism: a dennettian theory of existence," in *Dennett's Philosophy*, eds A. Brook and D. Thompson (Cambridge, MA: The MIT Press), 147–168. doi: 10.7551/mitpress/2335.003.0010

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philos. Psychol.* 22, 1–19. doi: 10.1080/09515080802703620

Schöner, G., and Kelso, J. A. S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science* 4847, 1513–1520. doi: 10.1126/science.3281253

Simons, P. (2000). *Parts: A Study in Ontology*. Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780199241460.001.0001

Suñé, A., and Martínez, M. (2019). Real patterns and indispensability. *Synthese* 19, 1–16. doi: 10.1007/s11229-019-02343-1

van Inwagen, P. (1995). *Material Beings*. Ithaca, NY: Cornell University Press, doi: 10.7591/9781501713033

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Vitanyi, P. (2009). Turing machine. *Scholarpedia* 4:6240. doi: 10.4249/scholarpedia.6240

Walter, S., and Kyselo, M. (2009). Fred adams, ken aizawa: the bounds of cognition. *Erkenntnis* 71, 277–281. doi: 10.1007/s10670-009-9161-2

Wilson, R. A. (2004). *Boundaries of The Mind: The Individual in the Fragile Sciences: Cognition*. Cambridge: Cambridge University Press.

Wilson, R. A. (2010). "Extended vision," in *Perception, Action and Consciousness*, eds N. Gangopadhyay, M. Madary, and F. Spicer (New York: Oxford University Press).

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

MINI REVIEW
published: 15 December 2020
doi: 10.3389/fpsyg.2020.538228

Determining the Function of Social Referencing: The Role of Familiarity and Situational Threat

Samantha Ehli[1], Julia Wolf[2], Albert Newen[2], Silvia Schneider[1] and Babett Voigt[1]*

[1]*Mental Health Research and Treatment Center (MHRTC), Faculty of Psychology, Ruhr-University Bochum, Bochum, Germany,* [2]*Department of Philosophy II, Ruhr-University Bochum, Bochum, Germany*

In ambiguous situations, infants have the tendency to gather information from a social interaction partner to regulate their behavior [social referencing (SR)]. There are two main competing theories concerning SR's function. According to social-cognitive information-seeking accounts, infants look at social interaction partners to gain information about the ambiguous situation. According to co-regulation accounts, infants look at social interaction partners to receive emotional support. This review provides an overview of the central developments in SR literature in the past years. We focus on the role of situational aspects such as familiarity of SR partners and situational threat, not only for SR (looking), but also for subsequent behavioral regulation (exploration, affect). As the competing accounts make different predictions concerning both contextual factors, this approach may reveal novel insights into the function of SR. Findings showed that a higher familiarity of SR partners consistently resulted in decreased looking (cf. social-cognitive accounts) and that higher threat remains largely understudied, but seemed to increase looking in the first few studies (cf. co-regulation accounts). Concerning behavioral regulation (exploration, affect) findings are mixed. We point out that moving toward a more complex situatedness may help to disentangle the heterogeneous results by considering the interaction between familiarity and threat rather than investigating the factors in isolation. From a general perspective, this review underlines the importance of situational factors and their interaction in eliciting a phenomenon, such as SR, but also in determining the nature of the phenomenon itself.

Keywords: social referencing, social-cognitive, information seeking, comfort seeking, co-regulation, infants, familiarity, situational threat, understanding others

OPEN ACCESS

Edited by:
Piotr Winkielman, University of California, San Diego, United States

Reviewed by:
Gabriela Markova, University of Vienna, Austria
Agnes M. Kovacs, Central European University, Hungary

****Correspondence:***
Samantha Ehli
samantha.ehli@rub.de

Specialty section:
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

Received: *20 July 2020*
Accepted: *17 November 2020*
Published: *15 December 2020*

Citation:
Ehli S, Wolf J, Newen A, Schneider S and Voigt B (2020) Determining the Function of Social Referencing: The Role of Familiarity and Situational Threat. Front. Psychol. 11:538228. doi: 10.3389/fpsyg.2020.538228

INTRODUCTION

Social referencing (SR) is the tendency of a subject (infant) to gather information from an informant (social interaction partner) in order to regulate one's behavior towards an ambiguous referent for which a fully accurate evaluation is missing (Zarbatany and Lamb, 1985; Walden and Kim, 2005; Striano et al., 2006; Stenberg, 2009; Fawcett and Liszkowski, 2015; Schieler et al., 2018). It emerges from the age of 7 to 10 months and forms a foundation for social learning and social appraisal in adulthood (Walle et al., 2017).

Despite a long tradition of SR research rooting back to the 1980s, there is an ongoing debate concerning the function of SR in infancy. In the classical social-cognitive view, infants refer to other persons in order to seek for information. This perspective is still the default to some extent today, but there are empirical challenges to this view. In 1996, Baldwin and Moses provided a seminal review of SR research in infancy. According to them, the empirical evidence for the classical social-cognitive view could also be fully explained by less demanding processes such as comfort seeking (co-regulation accounts). They recommended taking a situated perspective, that is, examining how the features of the referent and the features of the informant influence SR. Specifically, going beyond an individualistic cognitive approach, they called for research on two questions: How does the (1) familiarity of the SR partner and (2) situational threat influence SR?[1] As the accounts make different predictions about the influence of these two contextual conditions, the answer to these questions could provide critical novel insights into the function of SR, Baldwin and Moses (1996) argued.

In the past 24 years, several follow-up studies examined how the features of the informant and the referent affect SR. **Figure 1** briefly summarizes respective research. However, pursuing Baldwin and Moses (1996) idea, we specifically review research about the role of familiarity of the SR partner and situational threat and evaluate its implications for understanding SR's function. Mastering the ambiguous referent thereby means that children approach the ambiguous situation (exploration behavior) and/or that children express less negative affectivity (after referring to the informant). Thus, for conclusions about SR's function, the consideration of exploration behavior and affectivity is of critical importance (Carver and Vaccaro, 2007).

Before drawing conclusions regarding SR's function, we first describe the two SR-accounts and their predictions for the role of both contextual factors for SR and for infants' subsequent behavioral regulation (exploration of the referent and infants' affective expressions). Based on the example of these two contextual features, we will show that an increased sensitivity for the situatedness of SR is a key development in the field of SR. Finally, we discuss how a situated perspective may help disentangling whether a child's reason to refer to a SR-partner depend on the social and physical context.

THEORETICAL ACCOUNTS AND THEIR PREDICTIONS FOR THE INFLUENCE OF FAMILIARITY AND SITUATIONAL THREAT

Social-Cognitive Accounts

According to social-cognitive accounts, SR refers to children's search for information from the SR partner in order to evaluate an ambiguous situation (also referred to as classic information-seeking or information-gathering accounts; Bandura, 1992). These accounts imply that even very young children understand others as sources of information; that is, infants actively seek

[1]Authors raise a third question concerning which modalities (facial, bodily, verbal) influence SR to elaborate on the intentionality behind SR. As the question of intentionality is not in the focus of the present review, we do not address literature on the influence of modalities here.

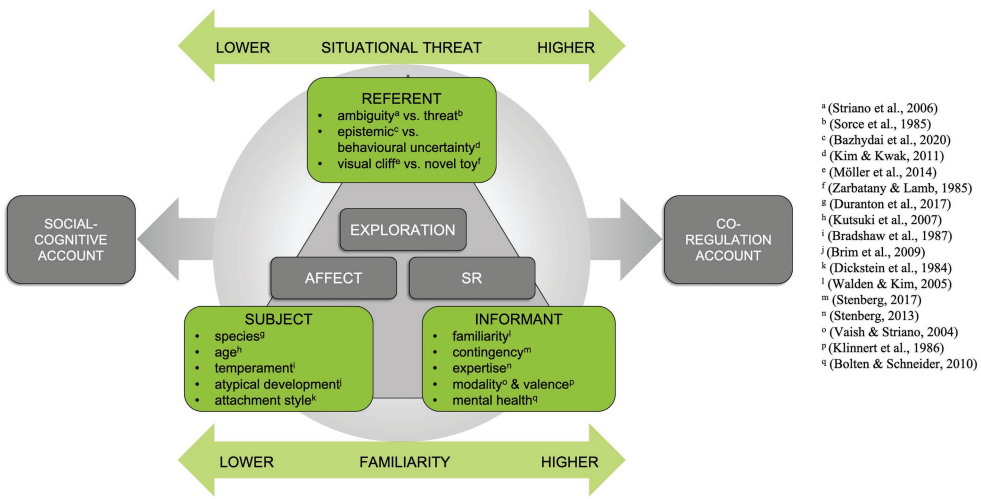

FIGURE 1 | The interplay among features of subject, informant and referent during social referencing (SR). The subject refers to the informant to gather information about the referent (SR). The informant's reactions influence subject's affect and the exploration of the referent. Several features of the subject, informant, and the referent have already been examined or are under suspicion to influence SR. The present mini-review focuses on the role of situational threat and familiarity to find out under which circumstances infants refer to the informant in order to gather information (social-cognitive account) or in order to receive emotional support (co-regulation account).

information before it is provided or even cause others to share their knowledge. Baldwin and Moses (1996) questioned this assumption given infant's poor performance in explicit theory of mind (ToM) tasks (Wellman et al., 2001). However, more recent findings suggest that infants pass implicit ToM tasks (see Scott, 2017 for a review). Further, evidence on pointing indicates that very young children understand ostensive gestures and use them to interrogate knowledgeable, but not ignorant, social partners (e.g., Liszkowski et al., 2008; Kovács et al., 2014). Thus, infants seem to possess the prerequisites for seeking information. This weakens Baldwin and Moses' hesitations towards social-cognitive accounts, which remains the most prominent explanation for SR in the current literature (e.g., Shaffer and Kipp, 2014; Meins, 2017).

Only recently, representatives of this theory considered social situational aspects such as familiarity of the SR partner. They predict that infants increase their looking toward more unfamiliar SR partners, as (a) they have a general preference for novel stimuli (novelty hypothesis, Roder et al., 2000), (b) they need more time to understand reactions of more unfamiliar SR partners (familiarity hypothesis, Stenberg, 2012), or (c) the experimenter is usually unfamiliar, but also more knowledgeable with regard to the laboratory context (expertise hypothesis, Feinman et al., 1992).

Such looking preference should lead to behavioral regulation (exploration) in accordance with the message of more unfamiliar SR partners, because the reactions of the preferred SR partner are more salient to the infant. However, consequences of familiarity for infants' affective expression are largely neglected by social-cognitive accounts and related studies (e.g., Striano and Rochat, 2000; Stenberg and Hagekull, 2007).

According to Baldwin and Moses (1996), making a context more threatening decreases its ambiguity, so that less information is needed to disambiguate the situation. Thus, social-cognitive accounts propose that SR and exploration should decrease with increasing threat. In the case of familiarity, they do not explicitly address how threat would affect infants' affectivity.

Co-regulation Accounts

Co-regulation accounts assume that children refer to social partners in order to seek comfort, to check for proximity, or to share affective experiences. These behaviors are not specific to ambiguous situations, but may also occur under these specific circumstances. For example, infants may refer to their mother as ambiguous situations usually elicit arousal, and infants have only limited skills to downregulate this arousal on their own (Kopp, 1989). Thus, SR is seen as one strategy for emotional regulation, in addition to seeking for physical proximity. From this perspective, SR bases on attachment processes (e.g., Ainsworth, 1992) and requires less advanced cognitive skills (Baldwin and Moses, 1996).

Familiarity plays a prominent role in co-regulation accounts. Familiar SR partners, particularly the mother, are seen as a secure base, which helps to maintain infants' arousal within an optimal range. Familiar interaction partners are usually more competent providers of emotional comfort as infants already learned to trust them and as familiar faces are easier to process (Stenberg and Hagekull, 2007; Ainsworth et al., 2015). In contrast to social-cognitive accounts, co-regulation accounts predict increased looking behavior to more familiar interaction partners and behavioral regulation in accordance with their reaction. This effect is not limited to primary caregivers but, if given the choice, children will generally prefer to look to more familiar SR partners. Further, threatening situations should increase children's arousal and their need for emotion regulation, resulting in increased SR, increased negative affect, and less exploration.

In short, both accounts consider situational factors, but make different predictions concerning the role of familiarity and threat (**Table 1**). In the next section, we review relevant findings to evaluate these predictions and their implications for the nature of SR.

TABLE 1 | Predictions for the influence of familiarity of the social interaction partner and situational threat on SR, exploration behavior and affectivity according to the social-cognitive accounts, and the co-regulation accounts.

	Social-cognitive accounts	Co-regulation accounts
Familiarity		
SR	unfamiliar > familiar	unfamiliar < familiar
Exploration behavior	in line with reactions of	in line with reactions of
Negative affectivity	unfamiliar informant	familiar informant
Potential threat (lower ambiguous threat vs. higher ambiguous threat)		
SR	lower > higher	lower < higher
Exploration behavior	lower > higher	lower > higher
Negative affectivity	lower = higher	lower < higher

Concerning familiarity, social-cognitive accounts propose that infants should increase looking towards a more unfamiliar person (novelty hypothesis/expertise hypothesis). As the behavior of this person becomes more salient for the infant, infants' behavioral regulation (exploration) should align with the reaction of the more unfamiliar person (e.g., more exploration, if the unfamiliar person provides a positive message). Co-regulation accounts propose that infants increase looking toward more familiar SR partners, resulting in behavioral regulation in accordance with more familiar SR partners' reactions to the referent (e.g., more exploration and less negative affect in case of a positive message). Concerning situational threat, social-cognitive accounts propose that SR and exploration should decrease with increasing threat (hence decreasing ambiguity), as less information is needed to disambiguate the situation. In contrast, co-regulation accounts propose increasing SR and negative affectivity and decreasing exploration with increasing threat, as there is a higher need for emotion regulation.

EMPIRICAL FINDINGS FOR THE INFLUENCE OF FAMILIARITY ON SR

The majority of research focuses on looking behavior as the core element of SR. Theoretical accounts of SR imply that children's search for information aims at dissolving the ambiguous situation. In empirical investigations, the ambiguous situation either refers to a novel toy (e.g., Mumme et al., 1996) or to a visual cliff (e.g., Striano et al., 2006). Extending the work of Baldwin and Moses (1996; who focused only on looking behavior), this review also takes into account SR's consequences for exploration and negative affectivity. All presented empirical evidence is based on data from children younger than 24 months.

Looking Behavior

Children generally increase their looking behavior toward other persons in an ambiguous situation to gather information (Carver and Vaccaro, 2007). As predicted by social-cognitive accounts, the majority of studies found that infants preferred to look at an unfamiliar experimenter compared to their looking behavior toward the mother (Walden and Kim, 2005; Stenberg and Hagekull, 2007; Stenberg, 2009; Kim and Kwak, 2011; Schieler et al., 2018). In only one exception that infants looked longer toward the familiar experimenter compared to an unfamiliar experimenter (Stenberg, 2012). Overall, these findings seem to support social-cognitive accounts. However, several concerns remain unresolved.

First, SR is only one of several strategies that infants use to overcome an ambiguous situation; seeking proximity is another. With unfamiliar SR partners the strategy of choice might be increased social looking, whereas with familiar partners it could be proximity seeking where children's looking pattern remains unaffected. Supporting this idea, Dickstein et al. (1984) found that social looking toward the mother decreases when proximity toward the mother increases, Ainsworth (1992) anecdotally described similar behavior in the strange situation task. Thus, physical proximity to the mother in the studies cited above may have biased the results toward social-cognitive accounts.

Second, it remains open whether familiarity or expertise explains the pattern in favor for social-cognitive accounts, as both features were conflated in most previous studies. Usually, the more experienced experimenter had more interaction time with the child or more speaking time. Evidence directly addressing expertise as an underlying factor for children's looking preference is mixed. In favor for the expertise account, Stenberg (2012, 2013) found that children preferred to look at the SR partner with more expertise if familiarity was kept constant. Another study attempted to examine familiarity and expertise further by testing some children in the laboratory and some at home (Schieler et al., 2018). In the laboratory, the experimenter might be considered the expert, while at home, the parent should have more expertise. Against the expertise hypothesis, Schieler et al. (2018) found increased looking toward the more unfamiliar experimenter in both contexts, even at home. Nonetheless, children might still have seen the experimenter as the expert, who instructed the parent (Walden and Kim, 2005). Hence, the question of familiarity vs. expertise as critical factor remains to be clarified by future studies.

Third, more fine-grained analyses of looking pattern data revealed that despite the preference for more unfamiliar interaction partners, infants increased looking behavior toward both the experimenter and the mother, when the former presented a novel toy (Schmitow and Stenberg, 2013). Infants seem to need reassurance from more familiar interaction partners to trust the information provided by unfamiliar, yet more knowledgeable SR partners. One interpretation may be that co-regulative and social-cognitive functions complement each other, a possibility that has not been tested empirically so far.

Exploration Behavior

While evidence of looking behavior seems to support social-cognitive accounts, the few findings relating to children's exploration behavior (of the ambiguous situation) are mixed. Stenberg and Hagekull (2007) found that children explored more with the unfamiliar experimenter than with the mother. Extending this evidence to other levels of familiarity, Schmitow and Stenberg (2013) found more exploration of a novel toy when it was presented by the unfamiliar experimenter as opposed to the familiar experimenter.

Analogous to looking behavior, expertise could explain the effect of familiarity in these studies (but see Zmyj et al., 2012, for contradictory findings in the context of imitation). Indeed, Stenberg (2013) showed that children increased their exploratory behavior more after receiving information from the expert experimenter. Here too the proximity to the mother (as a secure base and source for emotional comfort) may have biased the exploratory pattern in the direction predicted by social-cognitive accounts. However, both points cannot explain the results of studies that found the opposite pattern supporting co-regulative accounts. In those studies, children only approached the ambiguous situation after receiving information from their parent (Schieler et al., 2018) or a more familiar experimenter compared to a less familiar experimenter (Stenberg, 2012). Other studies even found contradictory findings, depending on which kind of exploratory behavior was analyzed. In Stenberg (2009), children looked more at a novel toy if the information was provided by the mother, but played more with it when the information came from the unfamiliar experimenter.

Taken together, it seems that infants show less (e.g., Schmitow and Stenberg, 2013), more (Schieler et al., 2018), or different explorative behavior (Stenberg, 2009) in the presence of their mother compared to an unfamiliar SR partner. However, when exploring familiarity independent of expertise, expertise seems to have the critical impact on the exploration behavior (Stenberg, 2012). Further, infants' behavior seems to be more affected by negative reactions of the social partner compared to positive ones (Vaish et al., 2008; Schieler et al., 2018), which may have obscured the influence of familiarity in some studies.

Thus, the current pattern for exploration behavior does not clearly speak in favor of one account. It must be borne in mind that the effect of SR on exploratory behavior is measured in much fewer studies, while measuring looking behavior is required for any SR paradigm. Hence, the heterogeneous findings result from a weak empirical base and await clarification in future studies.

Affect

In the context of SR research, affectivity could reflect an adequate emotional reaction to the ambiguous situation after receiving information about it (social-cognitive accounts). Alternatively, maybe emotional displays reflect the result of emotion regulation (co-regulation accounts).

The co-regulative pattern of lower negative affectivity in the presence of more familiar interaction partners receives little empirical support. Most studies found no significant differences in affect in the presence of SR partners of different familiarity (Walden and Kim, 2005; Carver and Vaccaro, 2007; Stenberg, 2009, 2012; Kim and Kwak, 2011). Usually, children showed relatively low levels of distress in any condition within the respective studies. Such low variability may explain the absence of effects on affectivity.

Overall, the findings about the influence of familiarity draw an inconclusive picture varying between and within domains (SR, exploration, and affectivity). Hence, whether SR's function aligns with the predictions of the social-cognitive or co-regulation account still remains open. Baldwin and Moses (1996) suggested a crucial role of situational threat as a second contextual factor, which could resolve the contradictory findings above.

EMPIRICAL EVIDENCE FOR THE INFLUENCE OF SITUATIONAL THREAT ON SR

Even though Baldwin and Moses already proposed in 1996 that situational threat might provide new insights on SR's function, only little progress has been made in this regard. In the few available studies, infants showed higher SR, less exploration (crossed a visual cliff less often, Striano et al., 2006), and increased negative affect (higher levels of arousal, Schwartz et al., 1973) on a steeper cliff (i.e., more threatening, less ambiguous) in comparison to a flatter cliff (i.e., less threatening, more ambiguous). Striano and Rochat (2000) found the same effect in a novel toy paradigm where they used a toy dog and measured infants' SR before the dog barked (lower potential threat) and after the dog barked (higher potential threat). SR increased with increasing threat. This supports co-regulation accounts that assume children should generally increase SR as one method of comfort seeking in highly threatening contexts. Besides the potential threat of a referent, other possible features have been neglected in research. For example, it might be interesting to assess the differences resulted by visual cliff vs. novel toy paradigms, as the former seems to have direct implications for infants' behavior (**Figure 1**). Findings from both ambiguous tasks have been used interchangeably.

CONCLUSION AND OUTLOOK

Baldwin and Moses have been pioneers in suggesting a stronger situatedness in investigating SR. This has led to a new direction in SR research, and respective findings give rise to new questions. In this review, we summarized research about the influence of two situational factors on SR – namely familiarity and threat. Baldwin and Moses proposed that the examination of both contextual factors (independent of each other) could help to elucidate SR's function. We reviewed respective research of the past 24 years leading to three major findings. First, higher familiarity of an interaction partner consistently resulted in decreased looking in many studies (in line with social-cognitive accounts). Second, only few studies examined the impact of familiarity on infants' subsequent exploration and affectivity with contradictory results. Third, situational threat remains largely neglected in empirical research, but seemed to influence SR, exploration, and affectivity in the few available studies (in line with co-regulation accounts). Thus, the function of SR may be more complex than previously suggested.

To resolve this puzzle, we suggest extending Baldwin and Moses' ideas and moving on from a simple situatedness to a more complex situatedness. This means not only considering both contextual factors independently, but also addressing the impact of familiarity in situations of different levels of threat. Rethinking the predictions of both accounts from this perspective results in new hypotheses. Social-cognitive accounts predict less relevance for SR as information-seeking strategy if the situation becomes more threatening. Hence, the preference to look at less familiar social partners should become less evident with increasing threat. In turn, co-regulation accounts assume more relevance of SR (as emotion regulation strategy) if the situation becomes more threatening. Thus, the looking preference for more familiar interaction partners should become particularly apparent with increasing situational threat. In other words, SR may serve different functions, depending on the current context conditions: shifting from information-seeking in highly ambiguous, less threatening conditions to emotion regulation in ambiguous but more clearly threatening contexts (**Figure 1**). Thus, we suggest that the question is not whether SR serves a social-cognitive or co-regulative function, but rather under what circumstances which function prevails.

Research examining the interplay of both contextual factors is missing so far, but we propose that this is a key strategy for clarifying the inconclusive findings about the function of SR. On a conceptual level, respective evidence would (a) unify both so far competing accounts on a higher hierarchical level and (b) underline the importance of a situated perspective for understanding the complex context-dependent nature of well-known developmental phenomena such as SR. Research investigating additional contextual factors that might modulate the role of SR would be a second promising avenue.

AUTHOR CONTRIBUTIONS

SE, JW, AN, SSch, and BV contributed conception and structure of this mini-review. SE and JW compiled the draft of the manuscript. BV wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - Projektnummer GRK-2185/1 (DFG-Graduiertenkolleg Situated Cognition). This publication is funded by the DFG-Graduiertenkolleg "Situated Cognition," GRK-2185/1.

ACKNOWLEDGMENTS

We thank Kathrin Lucht-Roussel from the library of the Ruhr University Bochum, who supported us in the literature search. Additionally, we thank Helen Vollrath, Elmarie Venter and Matej Kohár for their helpful comments.

REFERENCES

Ainsworth, M. D. S. (1992). "A consideration of social referencing in the context of attachment theory and research" in *Social referencing and the social construction of reality in infancy*. ed. S. Feinman (Boston, MA: Springer), 349–367.

Ainsworth, M. D., Blehar, M. C., Waters, E., and Wall, S. N. (2015). *Patterns of attachment: A psychological study of the strange situation classic*. New York: Psychology Press.

Baldwin, D. A., and Moses, L. J. (1996). The ontogeny of social information gathering. *Child Dev.* 67, 1915–1939. doi: 10.1111/j.1467-8624.1996.tb01835.x

Bandura, A. (1992). "Social cognitive theory of social referencing" in *Social referencing and the social construction of reality in infancy*. ed. S. Feinman (Boston, MA: Springer), 175–208.

Bazhydai, M., Westermann, G., and Parise, E. (2020). "I don't know but I know who to ask": 12-month-olds actively seek information from knowledgeable adults. *Dev. Sci.* 23:e12938. doi: 10.1111/desc.12938

Bolten, M., and Schneider, S. (2010). Wie Babys vom Gesichtsausdruck der Mutter lernen: Eine experimentelle Untersuchung zur familialen Transmission von Ängsten. *Kindheit Und Entwicklung* 19, 4–11. doi: 10.1026/0942-5403/a000002

Bradshaw, D. L., Goldsmith, H. H., and Campos, J. J. (1987). Attachment, temperament, and social referencing: interrelationships among three domains of infant affective behavior. *Infant Behav. Dev.* 10, 223–231. doi: 10.1016/0163-6383(87)90036-1

Brim, D., Townsend, D. B., DeQuinzio, J. A., and Poulson, C. L. (2009). Analysis of social referencing skills among children with autism. *Res. Autism Spectr. Disord.* 3, 942–958. doi: 10.1016/j.rasd.2009.04.004

Carver, L. J., and Vaccaro, B. G. (2007). 12-month-old infants allocate increased neural resources to stimuli associated with negative adult emotion. *Dev. Psychol.* 43, 54–69. doi: 10.1037/0012-1649.43.1.54

Dickstein, S., Thompson, R. A., Estes, D., Malkin, C., and Lamb, M. E. (1984). Social referencing and the security of attachment. *Infant Behav. Dev.* 7, 507–516. doi: 10.1016/S0163-6383(84)80009-0

Duranton, C., Bedossa, T., and Gaunet, F. (2017). Do shelter dogs engage in social referencing with their caregiver in an approach paradigm? An exploratory study. *Appl. Anim. Behav. Sci.* 189, 57–65. doi: 10.1016/j.applanim.2017.01.009

Fawcett, C., and Liszkowski, U. (2015). "Social referencing during infancy and early childhood across cultures" in *International encyclopedia of the social & behavioral sciences. 2nd Edn. Vol. 22.* eds. N. J. Smelser, P. B. Baltes and J. D. Wright (Amsterdam: Elsevier).

Feinman, S., Roberts, D., Hsieh, K., and Sawyer, D. (1992). "A critical review of social referencing in infancy" in *Social referencing and the social construction of reality in infancy*. ed. S. Feinman (Boston, MA: Springer), 15–54.

Kim, G., and Kwak, K. (2011). Uncertainty matters: impact of stimulus ambiguity on infant social referencing. *Infant Child Dev.* 20, 449–463. doi: 10.1002/icd.708

Klinnert, M. D., Emde, R. N., Butterfield, P., and Campos, J. J. (1986). Social referencing. The infant's use of emotional signals from a friendly adult with mother present. *Dev. Psychol.* 22, 427–432. doi: 10.1037/0012-1649.22.4.427

Kopp, C. B. (1989). Regulation of distress and negative emotions: a developmental view. *Dev. Psychol.* 25, 343–354.

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., and Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy* 19, 543–557. doi: 10.1111/infa.12060

Kutsuki, A., Egami, S., Ogura, T., Nakagawa, K., Kuroki, M., and Itakura, S. (2007). Developmental changes of referential looks in 7-and 9-month-olds: a transition from dyadic to proto-referential looks. *Psychologia* 50, 319–329. doi: 10.2117/psysoc.2007.319

Liszkowski, U., Carpenter, M., and Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition* 108, 732–739. doi: 10.1016/j.cognition.2008.06.013

Meins, E. (2017). "Emotional development and attachment relationships" in *An introduction to developmental psychology. 3rd Edn. Vol. 53.* eds. A. Slater and G. Bremner (Hoboken: BPS Blackwell).

Möller, E. L., Majdanddzic, M., and Bögels, S. M. (2014). Fathers' versus mothers' social referencing signals in relation to infant anxiety and avoidance: a visual cliff experiment. *Dev. Sci.* 17, 1012–1028. doi: 10.1111/desc.12194

Mumme, D. L., Fernald, A., and Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Dev.* 67, 3219–3237. doi: 10.1111/j.1467-8624.1996.tb01910.x

Roder, B. J., Bushnell, E. W., and Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy* 1, 491–507. doi: 10.1207/S15327078IN0104_9

Schieler, A., Koenig, M., and Buttelmann, D. (2018). Fourteen-month-olds selectively search for and use information depending on the familiarity of the informant in both laboratory and home contexts. *J. Exp. Child Psychol.* 174, 112–129. doi: 10.1016/j.jecp.2018.05.010

Schmitow, C., and Stenberg, G. (2013). Social referencing in 10-month-old infants. *Eur. J. Dev. Psychol.* 10, 533–545. doi: 10.1080/17405629.2013.763473

Schwartz, A. N., Campos, J. J., and Baisel, E. J. Jr. (1973). The visual the cliff: deep and cardiac shallow months and behavioral responses and nine on sides at five of age. *J. Exp. Child Psychol.* 15, 86–99. doi: 10.1016/0022-0965(73)90133-1

Scott, R. M. (2017). The developmental origins of false-belief understanding. *Curr. Dir. Psychol. Sci.* 26, 68–74. doi: 10.1177/0963721416673174

Shaffer, D. R., and Kipp, K. (2014). "Cognitive development: Piaget's theory and Vygotsky's sociocultural viewpoint" in *Developmental psychology childhood and adolescence*. ed. J. Perkins (Boston: Jon-David Hague, Wadsworth Cengage Learning).

Sorce, J. F., Emde, R. N., Campos, J. J., and Klinnert, M. D. (1985). Maternal emotional signaling: its effect on the visual cliff behavior of 1-year-olds. *Dev. Psychol.* 21, 195–200. doi: 10.1037/0012-1649.21.1.195

Stenberg, G. (2009). Selectivity in infant social referencing. *Infancy* 14, 457–473. doi: 10.1080/15250000902994115

Stenberg, G. (2012). Why do infants look at and use positive information from some informants rather than others in ambiguous situations? *Infancy* 17, 642–671. doi: 10.1111/j.1532-7078.2011.00108.x

Stenberg, G. (2013). Do 12-month-old infants trust a competent adult? *Infancy* 18, 873–904. doi: 10.1111/infa.12011

Stenberg, G. (2017). Effects of adults' contingent responding on infants' behavior in ambiguous situations. *Infant Behav. Dev.* 49, 50–61. doi: 10.1016/j.infbeh.2017.07.001

Stenberg, G., and Hagekull, B. (2007). Infant looking behavior in ambiguous situations: social referencing or attachment behavior? *Infancy* 11, 111–129. doi: 10.1111/j.1532-7078.2007.tb00218.x

Striano, T., and Rochat, P. (2000). Emergence of selective social referencing in infancy. *Infancy* 1, 253–264. doi: 10.1207/S15327078in0102_7

Striano, T., Vaish, A., and Benigno, J. P. (2006). The meaning of infants' looks: information seeking and comfort seeking? *Br. J. Dev. Psychol.* 24, 615–630. doi: 10.1348/026151005X67566

Vaish, A., Grossmann, T., and Woodward, A. (2008). Not all emotions are created equal: the negativity Bias in social-emotional development. *Psychol. Bull.* 134, 383–403. doi: 10.1037/0033-2909.134.3.383

Vaish, A., and Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Dev. Sci.* 7, 261–269. doi: 10.1111/j.1467-7687.2004.00344.x

Walden, T., and Kim, G. (2005). Infants' social looking toward mothers and strangers. *Int. J. Behav. Dev.* 29, 356–360. doi: 10.1177/01650250500166824

Walle, E. A., Reschke, P. J., and Knothe, J. M. (2017). Social referencing: defining and delineating a basic process of emotion. *Emot. Rev.* 9, 245–252. doi: 10.1177/1754073916669594

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304

Zarbatany, L., and Lamb, M. E. (1985). Social referencing as a function of information source: mothers versus strangers. *Infant Behav. Dev.* 8, 25–33. doi: 10.1016/S0163-6383(85)80014-X

Zmyj, N., Aschersleben, G., Prinz, W., and Daum, M. (2012). The peer model advantage in infants' imitation of familiar gestures performed by differently aged models. *Front. Psychol.* 3:252. doi: 10.3389/fpsyg.2012.00252

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. One editor of this special issue is also co-author of this publication; this did not compromise the impartiality as we submitted a blinded manuscript and suggested a non-affiliated editor for this publication. JW and AN as well as SE, AN, SSch, and BV have collaborated on a publication within the past 2 years. AN is a PhD-supervisor of JW and SE; SS is supervising SE.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 21 December 2020
doi: 10.3389/fpsyg.2020.566385

Minds, Brains, and Capacities: Situated Cognition and Neo-Aristotelianism

*Hans-Johann Glock**

Department of Philosophy, Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, Zurich, Switzerland

This article compares situated cognition to contemporary Neo-Aristotelian approaches to the mind. The article distinguishes two components in this paradigm: an Aristotelian essentialism which is alien to situated cognition and a Wittgensteinian "capacity approach" to the mind which is not just congenial to it but provides important conceptual and argumentative resources in defending social cognition against orthodox cognitive (neuro-)science. It focuses on a central tenet of that orthodoxy. According to what I call "encephalocentrism," cognition is primarily or even exclusively a computational process occurring inside the brain. Neo-Aristotelians accuse this claim of committing a "homuncular" (Kenny) or "mereological fallacy" (Bennett and Hacker). The article explains why the label "fallacy" is misleading, reconstructs the argument to the effect that encephalocentric applications of psychological predicates to the brain and its parts amount to a category mistake, and defends this argument against objections by Dennett, Searle, and Figdor. At the same time it criticizes the Neo-Aristotelian denial that the brain is the organ of cognition. It ends by suggesting ways in which the capacity approach and situated cognition might be combined to provide a realistic and ecologically sound picture of cognition as a suite of powers that flesh-and-blood animals exercise within their physical and social environments.

Keywords: situated cognition, neo-aristotelianism, brain, mereological fallacy, capacities, criteria, Wittgenstein

OPEN ACCESS

Edited by:
Achim Stephan,
University of Osnabrück, Germany

Reviewed by:
Eugen Fischer,
University of East Anglia, United Kingdom
Brian Paul McLaughlin,
Rutgers, The State University of New Jersey, United States

***Correspondence:**
Hans-Johann Glock
glock@philos.uzh.ch

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *27 May 2020*
Accepted: *16 November 2020*
Published: *21 December 2020*

Citation:
Glock H-J (2020) Minds, Brains, and Capacities: Situated Cognition and Neo-Aristotelianism. Front. Psychol. 11:566385. doi: 10.3389/fpsyg.2020.566385

INTRODUCTION

Situated cognition constitutes a powerful trend in contemporary cognitive science. One of its pillars is a fresh approach to philosophical problems concerning the very nature of the mental. More specifically, situated cognition raises questions about the ontology of cognition. What are the subjects of cognitive properties, states, and processes? What is the proper locus of cognition? Is cognition confined to the brain, or is it situated in whole bodies, organisms, and perhaps their environments?

Orthodox cognitive science is representationalist and computationalist. It treats cognition as a matter of calculations performed on symbolic or sub-symbolic representations. Both the representations and the computations operating on them are supposed to be implemented *inside the brain*. Accordingly, this orthodoxy subscribes to what I call "encephalocentrism"[1]. By contrast, situated cognition is part of an anti-subjectivist paradigm shift which also operates under the label

[1] Adams and Aizawa (2008) use "contingent intracranialism" and Hohwy (2016) "neurocentrism" as labels for positions they defend.

"4E Cognition" (Newen et al., 2018b). Cognition is *embodied* in that it is not confined to the brain, but involves the whole subject. It is *embedded* in that it is essential to cognition that this embodied subject is situated in a physical and social environment. It is *enactive*, in that it is equally essential that the subject operates actively within its environment, even in allegedly passive processes like perception. It is *extended* in that cognition may reach beyond the limits of the body, to features of the environment which are employed in understanding and explanation.

The impetus for this article is provided by the fact that there are venerable ancestors to situated-cum-4E cognition. More importantly, these ancestors have spawned contemporary work that is in many respects congenial to situated cognition, yet these parallels have so far gone largely unnoted[2]. Most importantly, attending to both the convergences and the differences sheds important light on the nature of cognition, and it holds the promise of overcoming encephalocentric opposition to situated cognition.

The ancestors and cousins of situated cognition that I have in mind are Aristotelian and/or Wittgensteinian currents within the philosophy of mind. To simplify matters, I shall henceforth speak of "Neo-Aristotelianism." Whereas, Wittgenstein himself revived the Aristotelian-cum-Thomist tradition unwittingly, others (Ryle, Anscombe, Geach) did so knowingly. Having been sidelined by the representationalist and computationalist mainstream since the 1960s, their perspective has been rehabilitated through the rediscovery of dispositions and abilities (Kenny, 1989; Hacker, 2007; Vetter, 2015; Schellenberg, 2018).

Neo-Aristotelianism revolves around a "capacity approach": a mind is neither a physical nor a mental substance, but a set of abilities which can be attributed and understood from a third-person perspective. This general parallel to the anti-subjectivist stance of situated cognition is supplemented by more specific parallels concerning the ontology of cognition. According to Neo-Aristotelianism it is neither a non-material soul nor a part of the body that cognizes—feels, desires, perceives, conceptualizes, thinks, infers, etc. Instead, it a *whole flesh-and-blood animal*—human or non-human—operating in its physical and social surroundings.

In this context, Neo-Aristotelianism challenges major tenets of the encephalocentrist orthodoxy. Encephalocentrism can take various forms. We must distinguish the claim that the brain and its parts are *subjects* of cognition—the things which possess cognitive properties, are in cognitive states or undergo cognitive processes—from the claim that they are the *location* of cognition. Since neither properties nor states have a spatial location, that second claim should be restricted to cognitive processes and activities. Next, one might distinguish *homuncular* from *non-homuncular* encephalocentrism. According to "homuncular functionalism," there are human-like agents in the brain that perform acts of cognizing such as computation and inference [see Lycan (1991)] and section Behavior and the brain below). But even according to non-homuncular functionalism, the property of undergoing a mental process/being in a mental state is the property of undergoing a neurophysiological process/being in a neurophysiological state. Since the subject of such neurophysiological processes and states is the brain or one of its parts, this implies that the latter is also the subject of mental processes and states. Finally, encephalocentrism can be more or less pronounced. A *strong* version has it that only the brain and its parts are subjects of cognitive states, processes, and activities; and all cognition takes place within the skull. Thus, according to classic functionalism, the cognitive system interacts with the environment, yet cognition is nonetheless completely realized by computational processes in the brain. A *weaker* version maintains only that the brain and its parts are among the subjects of cognition, and that some cognitive processes take place within the brain. Neo-Aristotelians take issue with all of these positions. Any ascription of cognitive states or processes to the brain is found guilty of a "homunculus fallacy" (Kenny, 1984, ch. 9) or a "mereological fallacy" (Bennett and Hacker, 2003; Maslin, 2006). In response, Dennett (2007) and Searle (2007) have defended weak encephalocentrism. Recently, there has also been a more radical reaction, according to which even demanding cognitive concepts apply literally to just about any biological phenomenon, basic physiological components of organisms included (Figdor, 2018).

If encephalocentrism could be vindicated in the face of Neo-Aristotelian criticism, the path to situated cognition would be blocked. For even weak encephalocentrism explains the cognitive powers and performances of whole animals by reference to *cognitive* powers and performances of their neurophysiological components. Accordingly, the *real subjects* of the fundamental cognitive states and processes are parts of individual organisms, and the *ultimate location* of cognition is within our skulls[3]. This would imply that situated cognition is at most a *derived* phenomenon. While it is the animal as a whole which is situated and operates within a material and/or social context, its cognitive exploits and abilities would be fully explicable by reference to its parts (organs). Even if the rest of the body and the activities of the whole animal within its environment must be taken into account, they have a bearing only through their impact on the cognitive phenomena in the brain. Their role can be accommodated within the encephalocentric orthodoxy[4].

[2]The exceptions concern Wittgenstein (Hutto and Myin, 2013; Kiverstein and Rietveld, 2015). There is as yet no comparison of situated cognition and Neo-Aristotelianism, and no discussion of their shared opposition to encephalocentrism.

[3]Thus Hohwy (2016) argues that (i) predictive processing, the view that "the brain minimizes its prediction error and thereby infers the states of the world" is "rapidly gaining momentum"; (ii) predictive processing implies a "neurocentrically skull-bound" picture of the mind; (iii) such a picture is incompatible with extended and embodied cognition. (i) is an understatement [one of many leading empirical neuroscientists relying on this paradigm is Dehaene (2020, ch. 2)]. (iii) is undeniable. And it is difficult to avoid (ii), unless one interprets talk of prediction and inferences as metaphors for information processing that causally enables cognitive processes without itself being cognitive. Consequently, for situated cognition there is a premium on showing that literal interpretations yield indefensible claims.

[4]Such accommodations include "extended functionalism" (Wheeler, 2010), theories incorporating environmsental factors into predictive processing (Clark, 2016), and Searle's view that we are "embodied brains" (2007, 119–21).

My contribution propounds a critique of encephalocentrism that sets store by the capacity approach while relinquishing other aspects of Neo-Aristotelianism. It starts out by indicating why an open-minded Wittgensteinian approach is preferable to Aristotelian essentialism, especially when it comes to linking up with situated cognition. Next, it argues that the labels "homuncular fallacy" and "mereological fallacy" are inaccurate, since the fundamental bone of contention is whether attributing cognitive and epistemic processes and abilities to organs and their parts amounts to a category mistake. On that issue, I side with the "Nonsense View" of Bennett and Hacker. I reconstruct the argument behind it and defend modified versions of its premises against animadversions by Searle and Dennett. Next I rebut Figdor's frontal attack on the Nonsense View. She appeals to current versions of encephalocentrism, such as predictive processing. I criticize her philosophical case for holding that these positions vindicate a literal interpretation of attributing cognition to the brain; yet it is not my ambition to demonstrate that they could only make a coherent contribution to neuroscience if they could be taken literally[5]. Instead, the next section concludes the philosophical dialectic: denying that the brain is the organ of cognition takes opposition to encephalocentrism too far. I end by summing up how the capacity view and situated cognition can benefit from each other.

THE NEO-ARISTOTELIAN FRAMEWORK

Neo-Aristotelians engage in a priori reflections of a metaphysical or conceptual kind. Situated cognition, by contrast, presents itself as thoroughly empirical and hostile to "armchair philosophy." It addresses the same questions concerning the nature and locus of cognition, "what" cognition is and "where" it takes place. But among its champions "there is a general agreement that a priori definitions and models of cognition are not helpful, and that we need to conduct experiments and consult the empirical literature" (Newen et al., 2018a, 9).

Fortunately, these two options are neither exhaustive nor incompatible. Empirical investigations no less than philosophical reflections rely on at least a preliminary understanding of what topic is being investigated. And we identify these topics through our established concepts, whether everyday, scientific or philosophical. These concepts are presupposed explicitly or implicitly not just by philosophical theories and arguments, but also by research projects, methods, and findings within the special sciences (see section Technical Uses and Metaphor).

Psychological notions are notorious for giving rise to a whole raft of vexatious puzzles. Prominent among them are the mind-body problem, the "riddle of consciousness," and the mark and scope of the cognitive, which is our topic. Any sober approach even to scientific problems concerning the mind should therefore pay attention to the established employment of the relevant expressions within their normal surroundings. Without the propaedeutic of conceptual clarification, we shall be "incapable of discussing the matter in any useful way because we have no stable handle on our subject matter" (Joyce, 2006, 52). Furthermore, we shall be liable to fallacies and confusions because of illicitly oscillating between different senses of pertinent expressions.

At the same time, such conceptual clarifications must be sensitive to the way in which concepts are understood and operationalized in scientific research programs and to their modification in the face of novel observational and experimental data (see Glock, 2017 and section Conclusion below). By contrast to Aristotelian essentialism, the Wittgensteinian strand in Neo-Aristotelianism acknowledges that our concepts are untidy and subject to change. This attitude is hospitable to a key ambition of situated cognition, namely to construct novel conceptual and methodological frameworks for the empirical study of cognition. Wittgensteinianism is also at odds with a Neo-Aristotelian tendency to equate the mind with the intellect (Kenny, 1989, 20–5; Hacker, 2013, ch. 1). Instead, it treats "mind," "mental," and their cognates as family-resemblance concepts. The phenomena they signify are united not by a single common feature, but by a complex network of overlapping and criss-crossing similarities (Wittgenstein, 1953, §§66–7). Such an approach supports another prominent conviction among proponents of situated cognition: there is no single "mark of the mental" (Newen et al., 2018a, 7, 10).

WHAT CAPACITIES CAN DO FOR UNDERSTANDING COGNITION

Even if Neo-Aristotelianism goes wrong in seeking to identify an immutable essence of the mind, its *capacity approach* is both enlightening and congenial to situated cognition. In the wake of Descartes, the mainstream of Western philosophy has treated "mind," its equivalents and cognates as the label of a special kind of thing, whether it be a separate mental substance, as in dualism, or the brain, as in materialism. The starting point of the capacity approach is negative: the mind is not a *bona fide* thing of any kind, whether mental–a *res cogitans*—or material—the brain (see White (1972), 464–5). Nor is it a kind of stuff or matter like water or gold: "mind" is a count-noun, and hence unsuitable as the name of a stuff.

The capacity approach also offers an alternative, by regarding the mind as a potentiality or power (Hacker, 2007, 90–121; Kenny, 2010). Potential properties are *bona fid*e attributes of particulars and substances, contrary to various forms of reductionism. At the same time, a potentiality must not be reified, treated as a thing of a peculiar kind that somehow co-exists with the particular or substance that possesses it. A power is neither a fiction, nor a flimsy actuality, nor an ethereal substance.

The central lesson: whether a subject has mental properties depends on what she is capable of doing. And to have a mind is to have a range of cognitive, volitional, and affective capacities or abilities. These must not be confused with

1. their exercise (in bringing about or undergoing physical or mental change);
2. the conditions that must obtain for manifesting or exercising the ability:
   - opportunity conditions: I may be able to dissect an angle with compass and ruler, yet lack the necessary equipment.

[5]For a discussion whether debates about the locus of cognition make a difference to *experimental* cognitive science see Kiverstein (2018, 23–4).

- enabling conditions: I may possess that ability and have the prerequisites but be impeded by disease (e.g., high fever) or injury (e.g., fractured hands).

3. their possessor (the individual animal or person);
4. their vehicle, that is the physical ingredient or structure of the possessor that sustains the ability, i.e., causally enables the possessor to exercise it (subject to opportunity and enabling conditions).

Judged from this perspective, dualism reifies mental powers, behaviorism reduces these powers to their exercise, and the mind-brain identity theory reduces them to their vehicle–the brain. Neo-Aristotelians have also relied on the capacity approach to resist encephalocentrism more generally.

FALLACIES AND CATEGORY MISTAKES

There is a venerable tradition of attributing mental and epistemic properties to things other than flesh-and-blood animals. Dualists like Plato and Descartes identify the mind with a non-material substance. For them, it is the soul that thinks (etc.) and thinking occurs "in" the soul. Against Cartesian dualism Wittgenstein insisted:

> Only of a human being and what resembles (behaves like) a human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious (1953, §281).

Alluding to this quote, Kenny criticized the "reckless application of human-being predicates to insufficiently human-like objects," notably the brain and its parts, in psychology and psycho-linguistics. He labeled this mistake the "homunculus fallacy" (Kenny, 1984, 125). For crediting the brain with perceiving, remembering, understanding, inferring etc. invites the question of how a conglomeration of neurons could do so. Since no convincing answer is forthcoming, one is driven to the absurd conclusion that there are homunculi in the brain who are capable of such cognitive feats.

Bennett and Hacker were in turn inspired by Kenny. But they also invoke "mereology," "the logic of part/whole relations." The "mereological fallacy" consists in "ascribing to a part of a creature attributes which logically can be ascribed only to the creature as a whole." It violates the "mereological principle": "psychological predicates which apply only to human beings or (other animals) as wholes cannot intelligibly be ascribed to their parts, such as the brain" (Bennett and Hacker, 2003, 29, 73). The Neo-Aristotelians concede that this mistake is

> not strictly speaking a fallacy, i.e., an invalid argument, since it is not an argument but an illicit predication. However, it leads to invalid inferences and arguments, and so can be loosely called a fallacy (Smit and Hacker, 2014, 1077; see also Kenny, 1984, 135-6; Bennett and Hacker, 2003, 73n13).

These terminological remarks call out for scrutiny. Note first that the difference between the alleged encephalocentric mistake and a fallacy in the "strict" logical sense does not just concern the level of complexity—a single statement (illicit predication) vs. a set of statements (invalid argument). In the case of a fallacy it is logically possible that the premises should all be true *and* the conclusion nonetheless false. But Bennett and Hacker are adamant that statements to the effect that parts of an animal cognize (perceive, experience, think, infer, etc.) are *nonsensical* rather than false (e.g., Bennett and Hacker, 2003, 2, 6, 72; Bennett and Hacker, 2007, 135). Unlike

(1) Mary cognizes

statements like

(2) Mary's brain / a part of Mary's brain cognizes

"make no literal sense." By these lights, statements like (2) cannot even be the *conclusion* of a fallacy. For they are bereft of linguistic meaning and hence *neither* true *nor* false.

Next, the Neo-Aristotelian defense of the tag "fallacy" invites a quick and dirty response. Many encephalocentrists concede that statements of form (2) are true not strictly speaking, but only in a loose or extended sense. According to Dennett (2007, 87–8), for instance, brains or their parts can only "sort of believe" (etc.). Bennett and Hacker object that in that case, a statement like (2) is only "sort of true," only "sort of explains" statements like (1), and only sort of makes sense (Bennett and Hacker, 2007, 140). By the same literalist standards, don't they inveigh against a "sort of fallacy"? The response is too quick, however. Bennett and Hacker's underlying complaint is that Dennett fails to explain what it is for parts of an animal to "sort of believe." By contrast, Neo-Aristotelians explain, albeit briefly, in what "extended sense" the mistake is a fallacy: it *leads* to fallacies.

Now, their opponents could retort that they can offer an analogous explanation:

(2*) Mary's brain / a part of Mary's brain *sort of* cognizes means

(2′) Neurophysiological processes in Mary lead to Mary's cognizing

However, the two cases do not run in parallel. In (2′), "lead" signifies a relation of mechanical causation. But when Neo-Aristotelians accuse encephalocentric predications like (2) of "leading" to logical fallacies, they mean an epistemic relation: (2) seems to vindicate an argument that is in fact invalid. Furthermore, in their story, one and the same subject commits both the illicit predication and the ensuing fallacy. By contrast, the *sub-personal* phenomena supposedly recorded in (2) and (2*) causally explain phenomena at the *personal* level recorded in (1), cognitive capacities and their exercises on the part of a whole animal.

But *what fallacy* is encouraged by applying psychological predicates to parts of an animal? Kenny's answer: if one explains cognitive phenomena at the personal level by reference to cognitive phenomena at the cerebral level, e.g., (1) by (2) or (2*), this invites the further conclusion:

(3) There are homunculi in Mary's brain that cognize.

But while the mistake is patent, the terminology is incongruent. The move from (2) to (3) is precisely *not* fallacious by Neo-Aristotelian standards. If psychological predicates can only be ascribed to human-like subjects, then from (2) something like

(3) *follows*. Kenny does not unmask a logical fallacy; he presents a *reductio ad absurdum* of encephalocentrism. (3) is absurd, irrespectively of whether it as an obvious empirical falsehood or conceptually incoherent.

Bennett and Hacker prefer the label "mereological fallacy" precisely because it does not impute a commitment to homunculi in the brain (2003, 73, fn.13). But their mereological take on the matter is also fraught with difficulties. They recognize that some psychological predicates can apply to parts of an animal, notably verbs of sensation, as in:

(4) Mary's hand hurts.

Accordingly, there is *no general principle* precluding the transfer even of psychological predicates from whole animals to their parts. Moreover, Bennett and Hacker also invoke Wittgenstein's afore-quoted dictum against ascribing psychological properties to objects that are *not* parts of animals, such as plants and computers. In conjunction, these two points show that the encephalocentric mistake cannot be a matter of mereology.

Encephalocentrist theories do not rely on general principles regarding the relations between wholes and their parts. Instead, many of them seem to be informed by an *inference to the best explanation*: As regards the specific case of *cognitive* activities, their being performed by a whole animal is best explained by there being a part of that animal which performs cognitive activities. In the eyes of encephalocentrists, empirical evidence demonstrates that statements like (2) or (2*) provide the best or perhaps even the only credible explanation of facts like (1)[6].

It is not that encephalocentric predications lead to fallacies. *Au contraire*. An inference to what encephalocentrists regard as the best explanation of cognition leads to applications of psychological predicates that Neo-Aristotelians regard as illicit. Ironically, this conforms to a sense of "fallacy" that differs from the logical one they employ: "a mistaken or delusory belief or idea, an error, esp. one *founded on unsound reasoning*" (OED, my emphasis).

The real allegation against encephalocentric predications is that they evince a *category mistake*. Ascribing cognition to the brain is not just unwarranted or false, it is bereft of sense. It applies mental predicates or concepts to things that are *not even potential candidates* for satisfying these concepts. The cognitive capacities and performances invoked in (2) can only be meaningfully attributed to the animal as a whole, and not—save metaphorically—to its parts[7].

We must keep this in mind when it comes to the justification of the Nonsense View. That view is not vindicated by the mereological principle. The latter is trivially true, since "apply only" is here intended as "are applicable intelligibly." It leaves open the contested issue: *why* should psychological predicates be meaningfully ascribable only to whole animals, not to their brains?

The argument behind the Nonsense View is best reconstructed as follows:

Criterial Premise: The ascription of a psychological predicate "F" to an object x is meaningful (and hence truth-apt) only if x can satisfy the criteria for being F.

Behavioral Premise: The criteria for an object x satisfying a psychological predicate "F" are human behavior or behavior resembling it on the part of x.

Wittgensteinian Conclusion: The ascription of a psychological predicate "F" to an object x is meaningful only if x can engage in human(-like) behavior.

Differentialist Premise: Neither the brain nor its parts can engage in human-like behavior.

Nonsense Conclusion: The ascription of a psychological predicate "F" to the brain or its parts is not meaningful (and hence not truth-apt)[8].

The argument is valid. However, all three premises require clarification and vindication, which will be provided in the next three sections.

PSYCHOLOGICAL CONCEPTS AND CRITERIA

In ordinary parlance, criteria are *ways of telling* whether something satisfies a predicate "F," and hence evidence for a claim of the form "x is F." That invites the suspicion that the Criterial Premise is merely "epistemological." It insists that there must be ways of finding out whether x satisfies F; but this has no bearing on the "ontological" issue of whether x is indeed F (Searle, 2007, 104–5). However, the requirement formulated in the Criterial Premise is *semantic*. In the Wittgensteinian employment of the term, the criteria for x being F are evidence of a particular type, "*logically* good evidence." Unlike inductive evidence, criteria are connected to x being F *conceptually* rather than through factual correlations established by experience. "F" would not mean what it does unless their fulfillment by x counted in favor of "x is F." "The criterial grounds of the ascription of a psychological predicate are partly constitutive of the meaning of that predicate" (Bennett and Hacker, 2003, 83).

Perhaps ontological questions are prior to epistemological ones. Yet semantic questions are prior to both, since matters of meaning antecede matters of knowledge *and* of fact. There can be true or false, justified or unjustified, answers to the question

[6]This holds for pioneers of cognitive neuroscience: "We seem driven to say that such neurons have knowledge" (Blakemore, 1977, 91); "If we are capable of knowing what is where in the world, our brains must somehow be capable of *representing* this information" (Marr, 1980, 3); "It is an inescapable conclusion that there must be a symbolic description in the brain of the outside world" (Frisby, 1980, 8). It is also a guiding theme in contemporary research. "[T]he theory that the brain is a sophisticated hypothesis-testing mechanism … is meant to explain perception and action and everything in between" (Hohwy, 2013, 2). Regarding an alternative theory according to which the brain uses "semantic pointers" to combine sensory, motor, and verbal presentations, Thagard claims that it is "the result of an inference to the best explanation of the full range of relevant evidence" (Thagard, 2019, 15).

[7]For the sake of argument, I assume that category mistakes are nonsensical rather than conceptually false. See Glock (2015).

[8]See Bennett and Hacker (2003, 71, 83). Their version of the Criterial premise runs: "The criterial grounds of the ascription of a psychological predicate are partly constitutive of the meaning of that predicate." My formulation eschews complications arising from their idiom of "criterial grounds" and "partly constitutive" (see section Criteria and Behavior) and from analytic functionalism, a position which accepts their premises while rejecting their conclusion.

"Is x F?"—"Does the brain cognize?"—only if that question is meaningful to begin with. That presupposes that the meaning of "F" has been determined at least provisionally. By the same token, there can be inductive evidence for x being F only if that condition is met.

Nevertheless, the semantics behind the Criterial Premise appears unduly verificationist[9]. Why should the meaningfulness of psychological predicates require criterial *evidence* that is available to us even in principle? But the Criterial Premise does not assume that we need to know *how to acquire evidence*, even under optimal conditions. It merely presupposes that it should be possible to *specify what such evidence would consist in*. Still, why can't one make do with specifying *application conditions* in the style of truth-conditional semantics? Now, such specifications can take various forms. One is disquotational, and follows the pattern "x satisfies the predicate "F" iff x is F". Applied to our case, this yields:

(5) The brain satisfies the predicate "cognizes" iff it cognizes.

Their popularity in formal semantics notwithstanding, however, statements like (5) do not properly *explain* the predicate quoted on the left-hand-side. They do not provide a *standard* for distinguishing correct and incorrect applications of "cognizes." And knowing statements like (5) is neither necessary nor sufficient for *understanding* "cognizes."

A second way of specifying application conditions is less vacuous.

(6) The brain satisfies the predicate "cognizes" iff it forms beliefs and desires on the basis of gathering and processing information

But in actual linguistic practice, even (6) would not count as an adequate explanation of "cognizes," if none of the explanantia on the right-hand sides could be somehow *operationalized* somewhere along the line. By the same token, someone who could only offer such explanations while being clueless about what kind of evidence *might count* for or against the fulfillment of the *explanantia* would at most be credited with a partial understanding[10].

Finally, *even if* the impossibility of specifying possible evidence for or against the satisfaction of a psychological predicate being satisfied were no bar to its being meaningful, it would deprive the predicate of any point in theories concerning the nature and causes of behavioral and mental phenomena that are even partly empirical. Encephalocentrism would be *hollow* as an approach in cognitive science.

Even in its philosophical manifestations, encephalocentrism aspires to making contributions to empirical theory formation. Unsurprisingly, therefore, it does not founder at the *general* semantic hurdle posed by the Criterial Premise. Encephalocentrism allows for evidence that the application conditions of psychological terms are fulfilled. Indeed, it positively specifies that this evidence includes data concerning neurophysiological goings-on. The real crux concerns the semantics of *mental expressions* in particular. What are the *pertinent criteria*, the criteria which give meaning to our psychological expressions?

CRITERIA AND BEHAVIOR

This question leads on to the Behavioral Premise. It rightly notes that in both everyday and scientific practice we ascribe psychological predicates to others on the basis of how they are disposed to behave. These are the criteria—the evidential grounds—for the fulfillment of these predicates. Nonetheless, in concluding that these evidential grounds are "partly constitutive" of the meaning of psychological predicates, Bennett and Hacker seem to "confuse the behavioral criteria for the *ascription* of psychological predicates with the *facts ascribed by these* psychological predicates" (Searle, 2007, 103). The application conditions for predicates like "x cognizes" and the truth conditions for statements like (1) concern the mind rather than behavior. By a similar token, to say "Mary is in pain" is not to say "Mary manifests pain behavior."

However, Bennett and Hacker deny explicitly that "the psychological predication is equivalent in meaning to the behavioral description the truth of which warrants its ascription (sic)" (2003, 82n35). "Criteria for the application of such a predicate are distinct from its truth-conditions—an animal may be in pain and not show it or exhibit pain behavior without being in pain" (2007, 210-11n18). The behavioral criteria for mental phenomena are "defeasible," subject to countervailing evidence (2003, 82–3).

Unfortunately, a puzzle remains. According to the Neo-Aristotelians, behavioral criteria are not just partly constitutive of the meaning of psychological predicates, they provide "constitutive grounds" for applying these predicates (Smit and Hacker, 2014, 1081). At the same time, they acknowledge that x can, for instance, be in pain without displaying pain behavior and that x can display pain behavior without being in pain. But in that case pain behavior does not *constitute* being in pain even partly, since pain behavior is not even a *necessary* condition for pain.

The resolution of the puzzle is to reconceive the conceptual relation between mind and behavior. First, behavioral criteria are not just defeasible but also diverse and context-sensitive. What counts as a manifestation of a mental state by one subject on one occasion, may not for another subject or another occasion. And what someone is disposed to do as a result of being in a particular mental state also depends on her other mental states (Geach, 1957, 8; Glock, 1996, 50–8, 93–7). Secondly, "constitutive grounds" are not facts that constitute the *phenomenon* of x being in a psychological state, but simply *non-inductive evidence* for x being in that state. At the same time, it is *constitutive of the meaning* of a psychological predicate that *there are*

[9]It is a "Wittgenstein-inspired rule-based semantics" (see section Elucidation vs. Revision). But it is not a "criterial semantics," since these rules need not specify criteria in the sense explained in section Criteria and Behavior. Still, they should specify conditions of application in an informative way capable of guiding linguistic practice.

[10]Consider a similar case: someone who can specify necessary and sufficient conditions for satisfying "tadpole"—being an amphibian at the larval stage of its life cycle—without being able to indicate what conceivable evidence (dis-)confirms something being an amphibian or larva is not a fully competent user.

behavioral patterns licensing its application independently of induction. Our psychological terms *would not mean what they do* if they were not bound up with *some* behavioral criteria or other, however diverse, context-dependent and defeasible. The capacity approach explains why this is so. Mental concepts have an essential connection to potentialities (dispositions, abilities). "Pain" would not mean what it does unless certain forms of behavior counted as manifesting pain in particular circumstances[11]. And it is part of psychological terms in general that they have some such manifestation. We would have no use for these expressions if they didn't[12].

This take on the Criterial and the Behavioral Premise suffices to support the Wittgensteinian Conclusion and to put paid to strong encephalocentrism. If, for example, we started to ascribe cognitive terms like "x perceives" or "x is intelligent" exclusively on neurophysiological grounds, in complete disregard of x's capacities to respond to its environment and to solve problems, these expressions would have changed their meaning. By the same token, although one can truthfully ascribe intelligence to a subject that does not manifest it, one can ascribe intelligence meaningfully—truly or falsely—only to a subject *capable* of behaving intelligently, a subject for which something counts as manifesting intelligence.

BEHAVIOR AND THE BRAIN

Repudiating strong encephalocentrism is compatible with ascribing mental properties to the brain and its parts *as well*. For it leaves open whether the Differentialist Premise holds. Can the brain and/or its parts behave in a way that satisfies the criteria of cognition? Dennett answers in the affirmative. He subscribes to the Wittgensteinian Conclusion of *Investigations* §281. At the same time, he denies that this precludes attributing mental attributes to neurophysiological phenomena. For

> brains and their parts *do* 'resemble a living human being (by behaving like a human being)'—and this resemblance is sufficient to warrant an adjusted use of psychological vocabulary to characterize that behavior (Dennett, 2007, 78).

Dennett admits that it would be illegitimate to attribute "*fully fledged*" mental phenomena to the brain parts (Dennett, 2007, 87). That would be to confuse the "personal" level of explanation which is "non-mechanical" with the "subpersonal" level which is "essentially mechanical" (Dennett, 2007, 78–9, 93). Nevertheless, one can attribute an "attenuated sort of belief and desire," stripped of many of their everyday connotations. "Just as a young child can *sort of* believe that her daddy is a doctor (without full comprehension of what a daddy or a doctor is), so a robot—or some part of a person's brain—can *sort of* believe that there is an open door a few feet ahead, or that there is something amiss over there to the right, and so forth" (Dennett, 2007, 87–8).

Dennett maintains that adopting such an "intentional stance" toward neurophysiological entities is a highly fruitful research programme that allows cognitive neuroscience to explain the foundations of our cognitive capacities. "Far from it being a *mistake* to attribute hemi-semi-demi-proto-quasi-pseudo intentionality to the mereological parts of persons, it is precisely the enabling move that lets us see how on earth to get whole wonderful persons out of brute mechanical parts" (Dennett, 2007, 88–9). This response faces two rejoinders. First, there is the need of explaining what *sort of* cognizing amounts to, not to mention "hemi-semi-demi-proto-quasi-pseudo intentionality." Dennett's allusion to the attenuated sense in which a small child can believe that her daddy is a doctor does not absolve him of that requirement. While the child cannot satisfy all of the criteria for holding such a belief, it can satisfy *some* of them (Bennett and Hacker, 2007, 141). Furthermore, she can fully satisfy criteria for believing *simpler things*, such as that there is a toy car in the room. Neither point holds of the brain or its parts.

Secondly, even if it made sense to credit sub-personal instances with cognition, wouldn't this only push further back the problem of explaining personal instances? One would then need to explain the representational capacities of these postulated homunculi, which engenders a vicious regress[13]. Now, the vacuity of explanations of human personal cognition by reference to sub-personal equivalents of human cognizers is acknowledged on all sides. That is why Dennett's "homuncular functionalism" invokes hierarchically structured "ever more stupid" intentional systems of a neurophysiological kind (Dennett, 1994, 240). Events at level E^n (cognition at the personal level) are explained by events at E^{n-1}, the latter by reference to events at E^{n-2}, etc. The aim is to discharge the explanatory task without embarking on an infinite regress, through a finite number of steps terminating in a level of completely non-intentional mechanisms.

Bennett and Hacker (2003, 141–7) recognize that there are levels of explanation between psychological descriptions of the whole animal and neuro-chemical descriptions of (parts of) the brain. They brook information-theoretic descriptions of (clusters of) neurons. They do not consider all the different notions of information currently on the market (see Adriaans, 2018). Nevertheless, they are right to distinguish "engineering information" from "semantic information." While the latter consists of true propositions that can be apprehended—believed, known—by an epistemic subject, the former concerns

[11] Our use of psychological expressions is not guided by exemplars. The grounds for ascribing pain to x is not that x *resembles*, e.g., the Man of Sorrows. It is that x *behaves* in a way that, in x's current circumstances, is a paradigmatic *manifestation* of pain. They resemble the features that characterize proto-/stereotypes in providing evidence that is defeasible. Flying will not count as evidence for x being a bird if x suckles its young or is invertebrate. But the defeating conditions for behavioral criteria are more context-dependent. Whether sobbing counts as evidence for x grieving for someone can depend not just on x's current setting and behavioral pattern but also on x's past history.

[12] We nevertheless understand ascriptions of mental states and processes to someone who is completely paralyzed. For the exercise of mental abilities can have behavioral manifestations only if certain enabling conditions obtain. I am grateful to a reviewer for this point, which indicates how the capacity approach can strengthen Wittgenstein's reflections.

[13] Unlike justification, explanation is *not* conditional. One can explain E_n by reference to E_{n-1}, *without* being able to adduce, even in principle, an explanans E_{n-2} (e.g. if E_{n-1} is the Big Bang). But that presupposes that events of *type* E^{n-1} are intelligible and uncontentious. According to the Nonsense View, purported explanations like (2) fail both conditions.

non-epistemic phenomena such as the probability of a datum and the freedom of choice in transmitting a signal.

There is a contrast between information as data or knowledge gained, possessed, and employed by whole animals on the one hand, mathematical constructs used to explain the causes and effects of neuro-chemical signals on the other. For this reason, the homuncular strategy does not address the crux of the debate: Is the application of psychological predicates (e.g., "possesses semantic information") to *anything other than whole subjects* starting with levels E^{n-1} conceptually licit in the first place?

If it is not, if E^{n-1} is amenable only to predicates like "processes engineering information", the attempt to causally explain phenomena at E^{n} through applying such predicates at E^{n-1} lacks sense and *a fortiori* explanatory power. The same holds for explaining (1) through (2) and (2*). By implication, explaining what (2) and (2*) *mean* by saying that they record the best causal *explanans* for (1) also fails. That Mary's brain cognizes—as of (2)—cannot mean that its cognizing causally leads to Mary cognizing, because there is no such thing as her brain cognizing. A related problem afflicts (2*). Either Mary's brain "sort of cognizing" is supposed to be a *genuinely cognitive and epistemic* episode; in that case we are facing the intelligibility question all over again. Or it is supposed to be an episode beneath that threshold, notably a neurophysiological or information-theoretic process; in that case, a causal explanation is on offer, yet it does not involve a contested encephalocentric predication.

A hierarchy of increasingly "dumb" homunculi raises questions about conceptual differences for each transition between levels of explanation. In consequence, the encephalocentrist faces a dilemma. *Either* he discharges the obligation to explain what *sort of* cognizing by parts of the brain amounts to through further mentalist and epistemic vocabulary. In that case we are back at square one, since it remains unclear what the application of such vocabulary to parts of the brain amounts to. *Or* he explains it by saying that it means that processes in the brain of a neurophysiological or information-theoretic kind causally explain the cognizing of the whole person. In that case the message is clear enough.

(2*)Mary's brain/a part of Mary's brain *sort of* cognizes

would amount to

(2#)Mary's brain/a part of Mary's brain undergoes a neurophysiological or information-theoretic process & that process is causally responsible for (enables) Mary's cognizing.

On that construal, (2*) involves a dual metonymical transfer, from a whole to its part, and from an effect to its cause. But then the attribution of mental properties to sub-personal instances is merely a figure of speech; indeed, it is a dispensable shorthand. The only remnant of encephalocentrism is the contention that the brain is causally responsible for cognition (see section Is the Brain the Organ of Cognition?).

TECHNICAL USES AND METAPHOR

There are alternative ways in which cognition might be attributed to parts of the brain in an attenuated, non-literal way. The first is that the use is technical. In that case, we would be dealing with *polysemes* of psychological expressions. But it would not just be incumbent on neuroscientists to explain their technical uses; they would also have to keep these uses apart from non-technical ones. Now, our mental concepts as applied to whole flesh-and-blood subjects determine the primary topics of philosophy of mind and cognitive science. The fundamental questions concerning mind and cognition are phrased in *extant, non-modified* vocabulary; indeed, mental idiom is first and foremost part of everyday discourse. We want to know, e.g., whether animals or brains think or are conscious in our sense of these terms, not in a sense introduced by new-fangled theories. To be sure, cognitive science also discovers and conceptualizes novel phenomena. And in tackling the initial topics it likewise introduces new concepts. For instance, the explanation of perception must employ technical concepts from a variety of areas, ranging from behavioral psychology to biochemistry. Yet statements couched in everyday terms like "Maria *saw* that Frank had put on weight," "Sarah *listens* to the *Eroica,*" "One can *smell* the wild strawberries," "The sense of *taste* is not affected by old age," "In the Müller-Lyer illusion two lines of equal length *appear* to be of unequal length," etc., pick out the basic *phenomena that the science of perception seeks to explain*.

Small wonder, then, that in presenting, interpreting, and drawing conclusions from their empirical data concerning perception, cognitive scientists do not uniformly stick to technical terminology. Instead, they often employ everyday terms like "representation," "symbol," "map," "image," "information" or "language" in ways which *either* remain unexplained *or* illicitly combine their ordinary uses with technical ones with an entirely different semantic import.

Any verdict to this effect needs to be sustained through painstaking analysis of individual cases. This cannot be undertaken here[14]. A general moral can be drawn nonetheless. The *explicit introduction* and *consistent employment* of homonyms of established psychological terms may be liable to cause confusion, yet it is unexceptional in principle. By that very token, however, it not merely avoids category mistakes; like the metonymical transfer in (2#), it eschews any encephalocentrism that the Nonsense View or situated cognition would have to resist. If belief*, knowledge*, and information* are used on the basis of neurophysiological or information-theoretic criteria, they apply to the *explanantia*—the phenomena which explain, respectively, the formation of beliefs, the acquisition of knowledge and the possession of information. Yet they do *not* univocally apply to these un-asterisked *explananda*.

A second way of attenuating the sense of an expression is metaphor. Weak encephalocentrists try to assuage doubts by pleading that applications of cognitive terms to the brain and its parts are metaphorical (Blakemore, 1990; Searle, 2007, 112;

[14]That such oscillations occur is granted even by Figdor, see section Philosophers, Nobel Laureates, and Nonsense. A reviewer helpfully suggested that in psycho-linguistics and in priming experiments "X is inferred from Y" is often understood as "X is activated by Y." This causal gloss causes confusion when combined with the familiar logico-epistemic one. For instance, when subjects are primed to draw inferences in that sense, areas in the right hemisphere are activated; from this it is concluded that the right hand hemisphere contributes to drawing inferences, without noting the switch to a causal sense (e.g., Mirous and Beeman, 2012).

Dennett, 2013). Metaphors serve a substantial heuristic function. They draw attention to features of the phenomenon to which they are applied by highlighting similarities with other phenomena. That is why they were traditionally regarded as abbreviated comparisons, (rightly so, see Schroeder, 2004). Metaphors are invaluable for many purposes. Still, if they are to lead our thinking in fruitful directions, they must be recognized as such.

This has important implications for allegations that encephalocentrists commit category mistakes. On the one hand, they must be made out separately for each contested case. On the other hand, it does not suffice for the accused to plead that they mean certain expressions to be taken metaphorically. They face two further demands. First, they must specify the respect in which the neurophysiological subjects of cognitive predicates resemble the personal ones. Secondly, that resemblance must suffice for the purposes—explanatory or justificatory—which the allegedly metaphorical use is meant to fulfill.

A metaphor draws attention to certain aspects of a phenomenon. But it contributes to an explanation only to the extent to which that phenomenon shares relevant features with things to which the metaphorical term applies literally. The purpose of using metaphors in an explanatory capacity must be to *compare* the explanandum with phenomena belonging to the literal extension of the term. The potential explanation is that there is an analogy between the explanandum and these phenomena.

ANALOGIES

At this point metaphors trade on a third way of attenuating the sense of an expression, namely to extend it by way of *analogy*. Paradigmatic examples include the extension of hydrodynamic notions such as "current" to the theory of electricity. Attributions of cognition to the brain are often explicitly defended as appeals to analogy. The idea fuels Gregory's animadversions to "semantic inertia" (Gregory, 1987, 242–3) and Dennett's insistence that sub-personal processes in the brain are "*strikingly like*" personal cognitive processes (Dennett, 2007, 86). In response, Bennett and Hacker complain that the "application of psychological expressions to the brain is not part of a complex theory replete with functional, mathematical relationships expressible by means of quantifiable laws as are to be found in the theory of electricity" (Bennett and Hacker, 2003, 77). However, Figdor (2018, ch. 3) argues that recent analogical theories revolve around precise models of relationships at a sub-personal level, nonetheless characterized in cognitive terms. The "temporal difference model" of reinforcement learning and the "predictive coding hypothesis" explain cognitive capacities and processes by exploiting "quantitative analogies" between neurophysiological phenomena and personal cognition, employing mathematical models and equations (2018, 31).

Nevertheless, there is a contrast between these novel theories in cognitive science and explanations of mechanical terms in physics, like "force" or "work." The latter diverge, often radically, from the everyday understanding of these terms. Yet they do so across the board, and in a clear-cut and mathematically precise manner, one giving rise to quantifiable laws[15]. What is more, they are patently fruitful. Figdor maintains that the aforementioned quantitative models are "highly confirmed." But there is no consensus concerning their precise interpretation, predictive accuracy or fertility. Indeed, many cognitive and life scientists concur with Bennett and Hacker's condemnation of the mereological fallacy[16].

There are also problems of principle with the analogy defense, even as applied to such quantitative theories. First, when cognitive labels like "learning" and "prediction" are not just quantitatively regimented but also transferred to neurophysiological subjects, they change their meaning, just as, e.g., "current" changed its meaning when transferred from liquids to electrodynamic phenomena. Secondly, in what does the analogy between the categories of these theories and the personal mental ones consist in? To avoid a *petitio* in favor of encephalocentrism, the sub-personal processes would have to be described in uncontentious terms, which means in terms that are either neurophysiological or information-theoretic. In that case, however, the analogies are of a *purely formal* or *structural* kind: certain mathematical models apply equally to both. In other respects, the employment, in particular the combinatorial possibilities and the inferential patterns, of the sub-personal expressions is far remote from that of the personal ones. It makes no sense to speak of columns of brain cells as inferring, calculating, predicting or perceiving *while going for a walk* (or for the purpose of foraging and preparing food)—these are not activities neurons can engage in.

More importantly, the incongruity also holds for psychological contexts. Neither the brain, nor one of its parts, nor neural tissues nor individual neurons can act on beliefs in conjunction with desires or goals; they do not show surprise when a belief or prediction turns out to be false, they do not modify their beliefs on account of the deliverances of their senses; they cannot be distracted in their cogitations by perceptual inputs, nor can cogitative assumptions taint their perceptions ("cognitive penetration"); they are not overwhelmed by emotions when experiencing joyful or sad situations; they do not avow their beliefs; they do not communicate their predictions and consider them in the light of objections by others. In short, what the mathematical models are capable of capturing in a way that is semantically and methodologically controlled and potentially fruitful once more concerns the causal enabling conditions of cognition, not features they share with cognition at the personal level.

But what of encephalocentrists who are prepared to go the whole hog? They might insist that brains, their parts, strata of neurons, individual neurons, etc., engage in all these mental

[15]Figdor insists that laws are unnecessary for "respectable biological theories" (2018, 95). But the extent to which biology aspires to such laws is a bone of controversy (Ayala and Arp, 2010, Part I). Furthermore, laws had better be part of the "model-based extensions" of psychological concepts based on *quantitative analogies* which she invokes against the Nonsense View.

[16]Figdor (2018, 98–100) plays down this fact by maintaining that their agenda differs from that of the Nonsense View. Yet even if this were true, it would not show that their agreement is based on misunderstanding or that they accept encephalocentrism after all.

activities. Not, of course, in the open-air but—taking work from home to extremes—inside the skull and in a neuro-chemical medium. Thus, Figdor deliberately casts to the wind a distinction by predictive processing theorists. To signify a mismatch between a predicted signal in the brain and the actual input, they use the technical neologism "surprisal" rather than the everyday "surprise." They should use the latter in a literal sense, she avers; for in both the personal and the sub-personal case there is a "discrepancy between an expectation and an observed actuality" (Figdor, 2018, 56). However, this obviously begs the question by assuming that there is a clear-cut similarity between "expectation" and "observation" at the personal and the sub-personal level. One cannot explain the use of one terminology—psychological idiom–in an area in which it is obscure and contested (the brain) through the use of another terminology—such as the idiom of behavior—in the same area, if that application is equally obscure and contested.

Admittedly, neurophysiological entities can behave, in the sense of causing change. But this holds of inanimate substances as well. It does so precisely because not all activity is psychological or guided by cognition. Perhaps one can mathematically model neural activity accurately in ways formally analogous to cognitive processes like predicting and adjusting expectations. Yet this no more shows that neurons actually engage in such cognitive activities than the fact that one can model the movement of planets through Kepler's laws shows that the planets deliberately follow these laws.

This last comparison highlights a final challenge facing hard-boiled encephalocentrists. Structural analogies with either human cognition and activity in this comprehensive sense are not confined to animate systems; they can be detected across the physical world. Radical encephalocentrists must be prepared to ascribe cognition, plus all of the concepts connected to it, to things on grounds of similarities, however thin and abstract, with personal subjects of cognition. If that were legitimate, what could bar ascribing them to any physical phenomenon whatever? Radical encephalocentrism threatens to lapse into panpsychism[17].

PHILOSOPHERS, NOBEL LAUREATES, AND NONSENSE

According to Figdor, applications of all psychological concepts, however sophisticated, to animate subjects of any kind, however primitive, are not just legitimate, they are to be taken at face value. This semantic doctrine—"Literalism"—goes along with "Anti-Exceptionalism," according to which "the relevant scientific evidence shows that psychological capacities are possessed by a far wider range of kinds of entities than often assumed. Literalism claims that, in contexts standardly interpreted as fact-stating, uses of psychological predicates to ascribe capacities in this wider range are best interpreted as literal with sameness of reference. Anti-Exceptionalism is the metaphysical position that underwrites the claim of sameness of reference" (Figdor, 2018, 5–6).

From this perspective Figdor attacks the Nonsense View. She quotes Bennett and Hacker: "If a form of words makes no sense, then it won't express a truth." She then turns their modus ponens into a modus tollens: ascriptions of psychological predicates to the brain "are expressions of truths (or empirically statements), so the Nonsense view fails" (Figdor, 2018, 98). But this tactic presupposes that the contested predications make sense. Figdor intimates two interconnected arguments for this presupposition. One is that these predications differ from clear-cut cases of nonsense like semantic anomalies. The other is that philosophers are not entitled to condemn pronouncements by Nobel laureates as nonsensical.

Both arguments ignore a central aspect of the Nonsense View. It is based on the idea that there are different kinds of nonsense or conceptual incoherence. Not all of them are gibberish or semantic anomalies. Certain types of philosophically relevant nonsense result from failure to pay heed to subtle features of concepts in the course of complex lines of reasoning, often as a result of being misled by powerful pictures and intellectual pressures. We are dealing with "latent" rather than "patent" nonsense (Wittgenstein, 1953, §464). This holds especially of category mistakes. There is no reason why scientists, however accomplished, should be immune to such confusions, especially when it comes to spelling out the implications of neurophysiological data for the psychological phenomena to be explained. Conversely, there is some reason to believe that philosophers can acquire both the conceptual sensitivity and the dialectical acuity to rectify such confusions. Even if that hope were overly optimistic, the numerous contradictions, paradoxes and antinomies that have been derived from apparently innocent premises and solid empirical findings provide ample evidence that conceptual inconsistencies and category mistakes need not lie open to view. The Nonsense View has it that encephalocentrists fall prey to such far from trivial mistakes. This results in a type of nonsense, since it cannot be spelled out coherently what encephalocentric predications mean in the context of the encephalocentrists' own explanations and arguments. Even if linguistic nonsense were the wrong category for category mistakes, the charge that encephalocentrism commits such mistakes would remain damning; and it cannot be dismissed simply by noting that famous scientists are not in the habit of talking gibberish.

Figdor acknowledges that Bennett and Hacker are right in complaining that cognitive neuroscientists often cause "confusion" by playing "fast and loose" with psychological predicates, notably by "defining terms in orthodox behaviorist manner and then drawing inferences that presuppose a cognitive interpretation" (Figdor, 2018, 104, 96, 98). But for her such lapses are confined to "the public communication of neuroscience." Bennett and Hacker, she contends, take account of "works intended for popular audiences," "they do not engage with the relevant scientific literature." This allegation misfires in two respects. First, it is incompatible with another dig Figdor takes at Bennett and Hacker, namely that "their view entails that Nobel

[17] Figdor denies that she is committed to panpsychism (Figdor, 2018, 9–10). Alas, she does not even intimate how it is to be avoided. Unsurprisingly, given that the similarities that connect human cognition to, e.g., information processing by bacteria, are so cheap that it is exceedingly difficult to draw a line.

prize-winning neuroscientist are writing nonsense in papers that helped garner them the prize" (Figdor, 2018, 94). Leaving aside the whiff of an appeal to Nobel authority, *if* the points raised by Bennett and Hacker indeed concerned only popular writings by neuroscientists, they could not entail any conclusions about their scientific publications, least of all if the two genres were as remote from each other as Figdor has it. Secondly, Bennett and Hacker cite numerous articles and books aimed at specialists (e.g., Bennett and Hacker, 2003, 75–81; Bennett and Hacker, 2007, 154–6). This leaves the worry that their targets are "often .... somewhat dated in neuroscience terms" (Figdor, 2018, 91). But on the same page, she concedes that "some recent peer-reviewed work in cognitive neuroscience … involves similar usage (or misusage)." So Figdor's allegation that Bennett and Hacker miss the "forest [serious neuroscience] for some epistemically inconsequential bushes nearby [popular neuroscience]" (Figdor, 2018, 100) is itself off the mark.

ELUCIDATION VS. REVISION

Figdor is nevertheless right in noting that "Bennett and Hacker and I are writing at cross purposes" (Figdor, 2018, 94). The reason is not, however, that she engages with respectable science whereas they target an Aunt Sally by restricting themselves to popularizations. It is rather that they are concerned with our *extant* concepts, whereas she explicitly charts and promotes a process of "conceptual revision" (1). On occasion, she acknowledges the radical nature of the proposed conceptual change (e.g., Figdor, 2018, 29). But she also maintains: "the rules for psychological predicates *have* changed" (Figdor, 2018, 96). This may hold to some extent for their application to non-human animals and computers. But the explicit conviction that these predicates apply non-metaphorically to organs, plants, and micro-organisms has not become entrenched in either quotidian or scientific discourse. In any event, the Nonsense View explicitly addresses our psychological concepts before the revolution propagated by Figdor. More importantly still, the crux of the matter is whether the extension (whether *fait accompli* or envisaged) is indeed governed by rules that are both consistent and do not simply change the topic. In deploring conceptual backsliding in popular neuroscience Figdor acknowledges willy-nilly both that this demand is legitimate and that it is frequently violated.

Like Figdor's Literalism in general, her rejection of the Nonsense View is based on Anti-exceptionalism. It depends on a realist semantics according to which scientific discoveries inform us not just about the actual extensions of psychological expressions, but also about their intensions or meanings. The rules have changed, this semantics implies, in the direction of capturing the real essences of psychological phenomena. In the wake of Kripke and Putnam, this has been a dominant view, and scrutinizing it is beyond the current remit. However, realist semantics has been explicitly criticized by proponents of a Nonsense View, especially as regards psychological expressions (Hacker, 1996; Glock, 2017). It is at odds with the "Wittgenstein-inspired rule-based semantics" that underlies their argument. This undermines Figdor's verdict that the Nonsense view fails even on its own terms (Figdor, 2018, 96, 100).

Empirical discoveries can show that the extension of extant concepts is different from what we used to think. Scientific theory revision, especially of a revolutionary kind, can also motivate conceptual change, the modification or replacement of those concepts. But, in line with the Criterial Premise, these concepts are determined by the criteria by which we decide *whether* something belongs to the extension. Therefore, scientific discoveries cannot show that our "traditional anthropocentric *grounds* for establishing the proper extensions of psychological predicates" are incorrect (Figdor, 2018, 104; my emphasis). As Davidson reminded us, "Our concepts are ours" (Davidson, 1999, 19). They play a role in our cognition, serve our epistemic needs and interests, and are geared to our capacities. To that extent, our extant *mental concepts* are anthropocentric; yet they are none the worse for that! Moreover, it does *not* follow that it is anthropocentric to insist that these concepts preclude application to brains and their parts.

Finally, the Wittgensteinian semantics undergirding the Nonsense view is more congenial to situated cognition than realist semantics. Situated cognition treats psychological concepts as means of making sense of others and ourselves, rather than as metaphysical lasers that "carve nature at its joints," in Plato's striking phrase.

IS THE BRAIN THE ORGAN OF COGNITION?

In another area, Neo-Aristotelianism is congenial to situated cognition up to a point. It shows that most of our extant psychological terms apply literally to whole subjects rather than their parts. This removes the pressure to *locate* cognition within a subject's skull. Questions like "Where did A perceive X/recognize that p/decide to Φ?" are answered by locating A in her environment (Bennett and Hacker, 2003, 128; Bennett and Hacker, 2007, 142–3). But Neo-Aristotelians take the rejection of encephalocentrism one step further. They deny that the brain is the *organ* of cognition.

> The stomach can be said to be digesting food, but the brain cannot be said to be thinking. The stomach is the digestive organ, but the brain is no more an organ of thought than it is an organ of locomotion [Fn25: One needs a normally functioning brain to think or to walk, but one does not walk *with* one's brain. Nor does one think *with* it, any more than one sees or hears with it]. If one opens the stomach, one can see the digestion of the food going on there. But if one wants to see thinking going on, one should look at the (sic) *Le Penseur* …, not at his brain. All his brain can show is what goes on there *while he is thinking* (Bennett and Hacker, 2007, 143).
>
> The brain is not an organ with which we can do anything, though we cannot do anything without a brain (Smit and Hacker, 2014, 1082).

The first, and uncontested, point to note: the brain is an organ, though complex in anatomical and physiological terms. In biology, an organ is a group of tissues that performs certain

functions. Like many organs (skin, liver, sexual organs), the brain fulfills a variety of functions.

Secondly, Bennett and Hacker deny that enabling cognition is one of these functions (Bennett and Hacker, 2003, 152). But their argument at most shows that the brain's function is not to enable cognition *on the part of the brain*. Furthermore, the denial is at odds with their observation that "the brain … enables the animal to see a visible scene" (2003, 139). Finally, it ignores the biological fact that enabling cognition is a crucial contribution that the central nervous system makes both to the well-being of individual animals and the adaptive advantage that its emergence conferred in evolutionary history.

Thirdly, Bennett and Hacker (2007, 135) acknowledge that one cannot cognize without the brain. To them, this does not show that the brain is the organ of cognition. For one cannot run without the brain either, and no one would say that the brain is the organ of locomotion. There is a difference, however. Neurophysiological processes and the proper functioning of the brain are the *proximate* causal enabling conditions of cognition. By contrast, the brain's causal relation to the movement of our locomotive organs is distal, mediated by motoric nerves, sinews, and muscles. Therefore, acknowledging that the brain is the organ of cognition does not commit one to maintaining that it is the organ of locomotion. In the terminology of the capacity approach, it is to say that the brain is the *vehicle* of cognition. It is that physical component of an animal which is directly responsible for its possessing cognitive capacities and causally involved in the exercise of those capacities.

Fourthly, Bennett and Hacker bluntly deny that we do anything *with* our brains. To be sure, we do not have direct voluntary control over what happens in our brains the way in which—through neurophysiological mechanisms like proprioception and motor nerves—we have control over the movement of our limbs. But as they recognize, this holds of other organs like the stomach as well. And we *do* digest with our stomachs.

Fifthly, established parlance suggests something analogous for cognition and the brain. According to Smit and Hacker "Use your brain!" simply means "Think!." "It no more signifies that we think with our brains than "I love you with all my heart" signifies that we love with our heart" (Smit and Hacker, 2014, 1089). But we employ "Use your brain!" to signify "Think!" *because we assume* that it is your brain that must operate properly for you to think. By contrast, we do *not* assume that your heart plays a special proximate role in enabling your emotions. "My brain isn't working properly today" is *not* a metaphor. It is on a par with "My stomach isn't working properly today." Both allude to causal factors influencing the enabling conditions of, respectively, my intellectual and metabolic capacities. That is why there is nothing *conceptually* amiss with trying to improve one's intellectual performance through "cognitive doping," imbibing drugs with neurophysiological effects.

Sixthly, that the brain is the organ of cognition is a major objection to 4E cognition (see Adams and Aizawa, 2008). Acknowledging this point and granting that there is a sense in which *we* think *with* our brains does not amount to backsliding into encephalocentrism. It no more entails that it is the brain that cognizes or that cognition occurs in the brain than the fact that our legs are the organs or vehicles of running entails that it is the legs that run on their own or that running occurs in our legs–as any marathon runner will testify.

Seventhly, Neo-Aristotelians are dead right that we cannot observe thinking in the brain. For one thing, the connection of cognition to neurophysiological phenomena is contingent, by contrast to its connection to behavioral capacities. For another, since cognizing is something done by whole subjects, it can only be observed by noting what these subjects do and are capable of doing. Nevertheless, in the brain we can observe neuro-chemical processes (indirectly, e.g., through fMRI scanners detecting rates of metabolism), and these processes do not merely accompany cognition, as Neo-Aristotelians have it, they *causally enable* it.

Eighthly, we can indeed observe digestion taking place in the stomach. Still, the contrast to cognition and the brain depends partly on how one conceives of digestion. At one level, digestion consists of chemical processes that take place in the gastrointestinal tract. But even a purely physiological account will have to include its interaction with other organs (liver, kidneys). In so far as digestion is the metabolic process that supplies energy to the whole organism, it is something that the whole organism engages in.

CONCLUSION

Situated cognition adopts such a wider, loosely speaking ecological, perspective. The capacity approach can provide a conceptual framework for this paradigm. Our psychological vocabulary captures *neither* neurophysiological or computational *nor* genetic-*cum*-genomic *nor* evolutionary differences, all of which are accessible at best on the basis of sophisticated instruments and theories. Instead, it captures differences in the kinds of *behavioral* and *perceptual* capacities human beings are both interested in and have unproblematic access to. This is unsurprising, especially from the perspective of situated cognition. We are social and cooperative primates by nature. Our languages include mental terms because of our fundamental need to describe, explain, predict and otherwise understand the behavior and behavioral dispositions of other human and non-human animals, and because of the equally fundamental need to express ourselves to other humans. No room here for the inner glow sought by Cartesians, or the neural mechanisms that captivate encephalocentrists.

Instead of emphasizing the brain at the expense of whole animals and their capacities, both situated cognition and the capacity approach adopt a perspective that is more realistic, and more naturalistic to boot. Cognitive and biological phenomena reveal themselves only when we go beyond the brain and consider not just the whole organism, but also the way the organism exercises its capacities in the context of its physical and social environment, in accordance with its "form of life," as Wittgenstein would put it.

Conversely, the capacity approach can profit from ideas and aspirations of situated cognition. Its chief merit lies in avoiding the misguided Cartesian riddle about how two ontologically

distinct substances like mind and body can causally interact, since it recognizes that the former is not a substance to begin with. Contrary to some advocates (Maslin, 2006, 209–19), however, it does not thereby dispatch the mind–body problem *tout court*. For one thing, capacities require a causal substratum, implementation or vehicle. This poses a scientific challenge—facing cognitive neuroscience and information theory—of explaining precisely how the vehicle of mental powers—the brain—causally sustains the power.

For another, capacities are defined by reference to how they are exercised. These exercises in turn are events and processes that stand in relations of efficient causation. Therefore, the question remains of what role causation plays for the episodic behavioral, mental and neurophysiological phenomena through which mental capacities are actualized or implemented. It won't do to claim, for instance, that feeling a pain is simply the actualization of the mental capacity for sentience. That answer is unexplanatory, not just in a factual-cum-scientific but also in a conceptual-cum-philosophical capacity.

At the scientific level, one needs to confront the question: what type of causal relation obtains between certain mental events and cognitive capacities on the one hand, neurophysiological processes, and structures on the other? More specifically, how do various mechanisms have to combine causally to constitute a suitable vehicle of cognition? What kinds of information processing need to occur in the central nervous system to provide its possessor with what kind of perception or intelligence? What neuro-chemical mechanisms can sustain such a flow of information? On such issues conceptual analysis should interact with empirical theory-formation of the kind undertaken within situated cognition. This article aimed to vindicate the shared opposition to encephalocentrism on which such interaction could be based.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This study was funded by University of Zurich, Faculty of Philosophy, Competitive Sabbatical Scheme.

ACKNOWLEDGMENTS

Work on this article was supported by a D-A-CH project The Nature and Development of our Understanding of Actions and Reasons (Swiss National Science Foundation Agreement #5100019E_177630) and by the NCCR Evolving Language Swiss National Science Foundation Agreement #51NF40_180888. I am grateful for comments by John Hyman, Severin Schroeder and Peter Schulte, and for assistance by Laura Burri, Basil M?ller, Cameron Alexander, Arthur Schwaninger, Christoph Pfisterer and Aneta Zuber.

REFERENCES

Adams, F., and Aizawa, K. (2008). *The Bounds of Cognition*. Malden: Blackwell.

Adriaans, P. (2018). *Information. Stanford Encyclopedia of Philosophy (Spring 2019 Edition)*, ed E. N. Zalta. Available online at: https://plato.stanford.edu/archives/spr2019/entries/information/ (accessed September 1, 2020).

Ayala, F. J., and Arp, R. (eds). (2010). *Contemporary Debates in Philosophy of Biology*. Oxford: Wiley-Blackwell, 141–164. doi: 10.1002/9781444314922

Bennett, M., and Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

Bennett, M., and Hacker, P. M. S. (2007). "The conceptual presuppositions of cognitive neuroscience: a reply to critics," in *Neuroscience and Philosophy*, eds M. Bennett, D. Dennett, P. M. S. Hacker, and J. Searle (New York, NY: Columbia University Press), 127–162.

Blakemore, C. (1977). *The Mechanics of the Mind*. Cambridge: Cambridge University Press.

Blakemore, C. (1990). "Understanding images in the brain," in *Images and Understanding*, eds H. Barlow, C. Blakemore, and M. Weston-Smith (Cambridge: Cambridge University Press), 257–283.

Clark, A. (2016). *Surfing Uncertainty*. Oxford: Oxford University Press.

Davidson, D. (1999). "Is truth a goal of inquiry?," in D. Davidson: Truth, Meaning and Knowledge, ed U. M. Zegleń (London: Routledge), 17–19. doi: 10.1017/CBO9780511625404.002

Dehaene, S. (2020). *How we Learn*. London: Penguin.

Dennett, D. (1994). "Dennett, Daniel," in *A Companion to the Philosophy of Mind*, ed S. Guttenplan (Oxford: Blackwell), 236–244.

Dennett, D. (2007). "Philosophy as Naive Anthropology: Comment on Bennett and Hacker," in Neuroscience and Philosophy: Brain, Mind, and Language, eds M. Bennett, D. Dennett, P. Hacker, and J. Searle (New York, NY: Columbia University Press), 73–95.

Dennett, D. (2013). Expecting ourselves to expect: the bayesian brain as a projector. *Behav. Brain Sci.* 36, 209–210. doi: 10.1017/S0140525X12002208

Figdor, C. (2018). *Pieces of Mind*. Oxford: Oxford University Press. doi: 10.1093/oso/9780198809524.001.0001

Frisby, J. P. (1980). *Seeing*. Oxford: Oxford University Press.

Geach, P. T. (1957). *Mental Acts*. London: Routledge and Kegan Paul.

Glock, H. J. (1996). *A Wittgenstein Dictionary*. Oxford: Blackwell. doi: 10.1111/b.9780631185376.1996.00002.x

Glock, H. J. (2015). Unintelligibility made Intelligible. *Erkenntnis* 80, 111–136. doi: 10.1007/s10670-014-9662-5

Glock, H. J. (2017). "Impure conceptual analysis," in *The Cambridge Companion to Philosophical Methodology*, eds S. Overgaard and G. d'Oro (Cambridge: Cambridge University Press), 83–107. doi: 10.1017/9781316344118.006

Gregory, R. (1987). "In defense of artificial intelligence," in *Mindwaves*, eds C. Blakemore and S. Greenfield (Oxford: Blackwell), 235–44.

Hacker, P. M. S. (1996). *Wittgenstein's Place in Twentieth Century Analytic Philosophy*. Oxford: Blackwell.

Hacker, P. M. S. (2007). *Human Nature*. Oxford: Wiley-Blackwell.

Hacker, P. M. S. (2013). *The Intellectual Powers*. Oxford: Wiley-Blackwell. doi: 10.1002/9781118609033

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001

Hohwy, J. (2016). The self-evidencing brain. *Noûs* 50, 259–285. doi: 10.1111/nous.12062

Hutto, D. D., and Myin, E. (2013). *Radicalizing Enactivism*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262018548.001.0001

Joyce, R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/2880.001.0001

Kenny, A. J. P. (1984). *The Legacy of Wittgenstein*. Oxford: Blackwell.

Kenny, A. J. P. (1989). *The Metaphysics of Mind*. Oxford: Oxford University Press.

Kenny, A. J. P. (2010). Concepts, brain and behaviour. *Grazer Philosophische Studien* 81, 105–113. doi: 10.1163/9789042030190_007

Kiverstein, J. (2018). "Extended cognition," in *The Oxford Handbook of 4E Cognition*, eds A. Newen, L. de Bruin, and S. Gallagher (Oxford: Oxford University Press), 19–40. doi: 10.1093/oxfordhb/9780198735410.013.2

Kiverstein, J., and Rietveld, E. (2015). The primacy of skilled intentionality. *Philosophia* 43, 701–721. doi: 10.1007/s11406-015-9645-z

Lycan, W. (1991). "Homuncular Functionalism meets PDP," in *Philosophy and Connectionist Theory*, eds W. Ramsey, S. Stitch, and D. Rumelhart (Hillsdale: Lawrence Erlbaum), 259–286.

Marr, D. (1980). *Vision*. San Francisco, CA: Freeman.

Maslin, K. T. (2006). *An Introduction to the Philosophy of Mind*. Oxford: Blackwell.

Mirous, H. J., and Beeman, M. (2012). "Bilateral processing and affect in creative language comprehension," in *The Handbook of the Neuropsychology of Language*, ed M. Faust (Oxford: Wiley-Blackwell), 319–341. doi: 10.1002/9781118432501.ch16

Newen, A., de Bruin, L., and Gallagher, S. (2018a). "4E cognition," in *The Oxford Handbook of 4E Cognition*, eds A. Newen, L. de Bruin, and S. Gallagher (Oxford: Oxford University Press), 3–15. doi: 10.1093/oxfordhb/9780198735410.013.1

Newen, A., de Bruin, L., and Gallagher, S. (2018b). *The Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780198735410.001.0001

Schellenberg, S. (2018). *The Unity of Perception*. New York, NY: Oxford University Press. doi: 10.1093/oso/9780198827702.001.0001

Schroeder, S. J. (2004). "Why Juliet is the sun," in Semantik und Ontologie, eds M. Siebel and M. Textor (Frankfurt am Main: Ontos Verlag), 63–101. doi: 10.1515/9783110327236.63

Searle, J. (2007). "Putting consciousness back in the brain," in *Neuroscience and Philosophy: Brain, Mind, and Language*, eds M. Bennett, D. Dennett, P. Hacker, and J. Searle (New York, NY: Columbia University Press), 97–124.

Smit, H., and Hacker, P. M. S. (2014). Seven misconceptions about the mereological fallacy. *Erkenntnis* 79, 1077–1097. doi: 10.1007/s10670-013-9594-5

Thagard, P. (2019). *Brain-Mind*. Oxford: Oxford University Press. doi: 10.1093/oso/9780190678715.001.0001

Vetter, B. (2015). *Potentiality*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198714316.001.0001

Wheeler, M. (2010). "In defence of extended functionalism," in *The Extended Mind*, ed R. Menary (Cambridge MA: MIT Press), 245–270.

White, A. (1972). Mind-brain analogies. *Can. J. Philos.* 1, 457–472. doi: 10.1080/00455091.1972.10716032

Wittgenstein, L. (1953). *Philosophical Investigations. 4th Edn*. Oxford: Wiley-Blackwell (2009).

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 14 January 2021
doi: 10.3389/fpsyg.2020.569403

Proprioception in Action: A Matter of Ecological and Social Interaction

Ximena González-Grandón[1,2,3]*, *Andrea Falcón-Cortés*[4] and *Gabriel Ramos-Fernández*[5,6]

[1] Departamento de Educación, Universidad Iberoamericana, Ciudad de México, Mexico, [2] Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, Mexico, [3] Insituto de Filosofía y Ciencias de la Complejidad IFICC-Chile, Santiago, Chile, [4] Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico, [5] Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico, [6] Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

The aim of this paper is to provide a theoretical and formal framework to understand how the proprioceptive and kinesthetic system learns about body position and possibilities for movement in ongoing action and interaction. Whereas most weak embodiment accounts of proprioception focus on positionalist descriptions or on its role as a source of parameters for internal motor control, we argue that these aspects are insufficient to understand how proprioception is integrated into an active organized system in continuous and dynamic interaction with the environment. Our strong embodiment thesis is that one of the main theoretical principles to understand proprioception, as a perceptual experience within concrete situations, is the coupling with kinesthesia and its relational constitution—self, ecological, and social. In our view, these aspects are underdeveloped in current accounts, and an enactive sensorimotor theory enriched with phenomenological descriptions may provide an alternative path toward explaining this skilled experience. Following O'Regan and Noë (2001) sensorimotor contingencies conceptualization, we introduce three distinct notions of proprioceptive kinesthetic-sensorimotor contingencies (PK-SMCs), which we describe conceptually and formally considering three varieties of perceptual experience in action: PK-SMCs-self, PK-SMCs-self-environment, and PK-SMC-self-other. As a proof of concept of our proposal, we developed a minimal PK model to discuss these elements in detail and show their explanatory value as important guides to understand the proprioceptive/kinesthetic system. Finally, we also highlight that there is an opportunity to develop enactive sensorimotor theory in new directions, creating a bridge between the varieties of experiences of oneself and learning skills.

Keywords: enactive cognition, sensorimotor theory, perception-action-coupling, ecological self, social cognition, agent-based models, kinesthetic phenomenology

OPEN ACCESS

Edited by:
Beate Krickel,
Technical University of Berlin, Germany

Reviewed by:
Elmarie Venter,
Ruhr University Bochum, Germany
Mateusz Wozniak,
Central European University, Hungary

***Correspondence:**
Ximena González-Grandón
glezgrandon@gmail.com

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *04 June 2020*
Accepted: *09 November 2020*
Published: *14 January 2021*

Citation:
González-Grandón X, Falcón-Cortés A and Ramos-Fernández G (2021) Proprioception in Action: A Matter of Ecological and Social Interaction. Front. Psychol. 11:569403. doi: 10.3389/fpsyg.2020.569403

INTRODUCTION

Suppose you have just woken up and immediately you feel the presence of your body; then, or maybe at the same time, you feel a body that is not yours cuddling you and perhaps also the sheets that do not cover your feet, leaving them uncovered. Your own body experience is subtly transformed with each focus of attention, as it takes on a distinctively ecological and social dimension. Both agents are sharing this proprioceptive and kinesthetic experience with each other. Can such embodied, ecological, and social interaction be part of an agent's proprioceptive perceptual experience?

In embodied accounts of proprioception, there are some disagreements about the explanatory role of the non-neural elements in real-time interaction. Alsmith and De Vignemont (2012, p. 1–13), for instance, propose a distinction between weak and strong approaches to body involvement. In the weak embodiment account, mental representations in bodily formats play a central role in action and perception, while moving bodies in interaction—a non-brain-bounded element—play a trivial one. These "B-formats" are associated with muscular sensation, as a physiological condition of the body (Goldman and de Vignemont, 2009), and become crucial when they are centrally represented in the brain and instantiated in internal models (Goldman, 2012). Strong embodiment accounts, in contrast, consider the whole body in its dynamical gestalt-like relations with physical and social environments as non-neural elements that play a strong causal and constitutive role in perception and action (Varela et al., 1991; O'Regan and Noë, 2001; Gallagher, 2017). Here, perception is a bodily experience intimately linked to skillful and effective embodied possibilities for action. Moreover, in these accounts, proprioception is better understood coupled with kinesthesia (as a proprioceptive-kinesthetic coupling or PK for short), a perceptual system that results from an active and ongoing coupling between feeling and performing.

Traditional accounts of proprioception place a special emphasis on the "self-perception" related to the body awareness of an agent's relative position in space. This is a central idea that can be found in weak embodiment approaches. Commonly, this positional sense description comes from Sherrington (1907) and his original conceptualization, in which the central nervous system (CNS) receives information about the spatial location of body parts and body segments to enable movement coordination. According to this, the experience of proprioception is described as a key source of spatial parameters for internal motor control at the level of the sensor: if an agent wants to put an earring into her earlobe, for example, she needs to wiggle her fingers around a bit to get it in and find the piercing hole. Here, a flexible transformation from proprioceptive afferent information about the position of the fingers is needed, for the capacity to estimate the appropriate set of motor commands required to achieve the desired outcome. In this model, however, experiencing one's body comes from verifying whether these estimations match or not in a controlled act, and the possibilities for variations are thus almost entirely determined.

The main objective of the present article is to introduce a strong embodiment account of proprioception based on O'Regan and Noë (2001) enactive sensorimotor theory of perception (ESMT) and Sheets-Johnstone (2019, 2020) kinesthetic phenomenology; as well as offer a formalization of this proposal following the work of Buhrmann et al. (2013) and Vicsek et al. (1995). This alternative account considers PK as a perceptual experience of spatiotemporal self-orientation in present action and interaction. On the one hand, from an enactive point of view—one that sees the perceiver as an active organism engaging with the ecological and social world—how the agent puts an earring into her earlobe depends on where her fingers are in relation to the rest of her body and where the piercing perforation is, how it feels, the previous experiences putting an earring, whether the surface where she is standing is flat or not, whether another agent is helping her, etc. This suggests that for the action to be effective, we not only need to perceive the objects on which we act or the state of the effector, such as the earring when inserted, but also the real-time PK experiences of the lived body whether dancing or walking.

In this view, a strong embodiment account of the proprioceptive perceptual experience should articulate, in operational and (if possible) formal terms, what these meaningful and skilled relations consist of. Here, we tackle this challenge by arguing that the PK perceptual experience is not only caused by some internal process in the brain—as a B-format representation or a specific somatosensorial cortex correlate—but rather that it is constituted by an organism's set of abilities to act during the ongoing affair of establishing meaningful relations with one's body and the world (O'Regan and Noë, 2001), that is, its proprioceptive-kinesthetic contingencies (PK-SMCs).

We propose that some dynamical self-oriented and relational features of the phenomenology of PK, resulting from coupling of perception and action, constitute the PK perceptual experience. Specifically, this is manifested in at least three different dimensions by the various degrees in which this experience occurs during a common episode of being present and bodily aware and ready to act: first, PK-SMCs-self that are related to the agent's own spatio-temporal self-orientation, in relation to other parts of one's body, and possibilities for action in present time; second, PK-self-ecological, which are those that arise from the agent's own embodied activity when interacting with the environment; and third, PK-self-other, which are those that arise from the agent's own activity when interacting with others. We will argue that these relational dimensions can be useful tools for explaining the PK-SMCs perceptual experience.

Finally, we illustrate the usefulness of these distinctions by applying them to the analysis of a model of minimal cognition of collective movement perception (following the work of Vicsek et al., 1995; Beer, 2003 and others). In this model, spatial and temporally organized behavior arise in agents with both skilled PK and non-skilled PK and in agents with any recourse to PK (deafferented agents) moving continuously inside a square. With this model, we achieve the dual purpose of testing the operational character of conceptual claims about PK perceptual experience from a strong embodiment account, and of bringing together ESMT and phenomenology while showing some limitations of the weak and current accounts.

WHAT IS PROPRIOCEPTIVE/KINESTHETIC COUPLING?

In order to have different opportunities of movement and to behave adequately in different environments, both known and unknown, an organism that recognizes itself separate from the environment has to master particular skills. The ability to recognize being in "the zero point of orientation" (Husserl, 1989) and being the origin of one's own movement, as a form of sensitivity to embodied actions, requires the concurring

development of the skills to experience the spatio-temporal self-orientation, and the feeling of possibilities for action. In this section, we will argue that proprioception and kinesthesia (as a PK coupling) have a central role in the development of this ability (Gallagher, 2003; Gapenne, 2014). In further sections, we will see that PK is also relevant to engage successfully in ecological and social interactions.

From a physiological standpoint, proprioception encompasses information from specialized sensory mechanoreceptors primarily found in muscles, such as neuromuscular spindles or neurotendinous organs, but also in the joints, tendons, ligaments, articulatory capsules, vestibular apparatus, or skin. These receptors transduce mechanical events into neural signals (Proske and Gandevia, 2012). In fact, muscle spindles provide the central nervous system (CNS) with afferent information about the length and velocity of the muscle in which the spindles are embedded and their rate of change, contributing to joint position sense and postural control. Traditionally, this has been considered as the main source of proprioceptive feedback for spinal sensorimotor regulation and servo-control (Sherrington, 1907; Fourneret and Jeannerod, 1998; Hewett et al., 2002)[1].

In this sense, proprioception is the perception of the relative positions of different body parts, where suitable proprioceptive sensors register joint angles and the activity of the effectors to which they are linked. These ideas are more aligned with the weak embodiment account. When trying to understand what the content of proprioceptive perceptual experience is, authors like Goldman (2012) or Goldman and de Vignemont (2009) have appealed to the existence of non-propositional B-formats. These are internal representations "associated with the physiological conditions of the body, such as pain, temperature, itch, muscular and visceral sensations, vasomotor activity, hunger, and thirst" (Goldman and de Vignemont, 2009, p. 156). Following these authors, B-formatted representations may originate peripherally and involve proprioceptive or kinesthetic information about the agent's own muscles. However, when represented centrally, they become genuinely B-formatted representations: "for example, codes associated with activations in somatosensory cortex and motor cortex" (Goldman, 2012, p. 74). When considering proprioception from this perspective, an implicit representationalist and brain-centered bias may emerge, where actual sensing and moving bodies play a marginal role. Indeed, this weak embodiment perspective restricts proprioception to the sensations about position produced by the static body and does not include the organization and the quality of the possibilities for movement from the proprioceptive self.

At this point, some accounts distinguish between proprioception and kinesthesia. For instance, human physiology has traditionally distinguished static sensations of one's joint positions (proprioception), from dynamic sensations, such as those that are sensitive to the rate of a specific movement (kinesthesia) (Kiefer et al., 2013). Indeed, kinesthesia was originally recognized as "the muscle sense," the sense of actions of the limbs (the sense of one's own movement), or the perceived sensations of positions in a system of possible movements (Sherrington, 1918). In this article, rather than subsume kinesthesia to proprioception or vice versa, or propose a distinction between them, we follow Sheets-Johnstone (2019) and Gapenne (2014) to suggest that proprioception is necessarily coupled with kinesthesia and possibilities for action (Gapenne, 2010, 2014): an emergent form of organization between sensing the spatio-temporal self-orientated body and the possibilities for the performing body.

Closer to the strong embodiment perspective, we argue that proprioception separated from kinesthesia fails to do justice to the different levels of analysis on which organisms' perceptual experience can be described. In the next section, we argue that this coupling can be understood more precisely in an ecological context.

Ecological Laws in PK

As argued by several investigations, although perception and action are mediated by different processes and pathways, they are coupled by ecological laws that relate afferent variables to parameters of the action system to regulate behavior adaptively (Varela et al., 1991; Warren, 2006; Dayan et al., 2007; Gonzalez-Grandón and Froese, 2018). This is implied by the notion of perception-action coupling from an ecological standpoint, which is made explicit by Gibson (1977, p. 223) in the following passage: "We must perceive in order to move, but we must also move in order to perceive." From this perspective, the perceptual prominence of vertebrate movement might come from these close interactions and regularities: the so-called ecological laws, such as attractors in the underlying dynamics between perception and action (Warren, 2006; without assuming predetermined or *a priori* cognitive or neural models; Dayan et al., 2007).

These ideas are a crucial background to the emergence of ESMT, an action-oriented perspective relying on enaction—putting into practice through action—where perceptual contingencies are intrinsically tied to specific movements. As Noë (2004, p. 2) states, perception is a "species of skillful bodily activity." In the coupling case we are concerned with, these ecological laws would be related to proprioception and kinesthesia. Proprioceptive information is both generated by and reciprocally used to regulate kinesthetic possibilities for movement. By information, Gibson (1977) meant spatio-temporal proprioceptive patterns of joint, muscle, or skin deformation at a moving limb, that are lawfully related to properties of the perturbations of the environment or aspects of the possibilities for the action itself. We can elaborate on this notion in terms of perception-action coupling.

An illustrative example comes from motor development in infancy, where researchers have begun to entertain that perceptual and motor systems develop in interdependent trajectories. Thelen (1990) provides evidence that motor skill

[1]A general description of proprioceptive feedback (PF) as an integral component of vertebrate locomotor control (Prochazka and Ellaway, 2012; Gordon et al., 2019) would be the result of two different processes: self-generated reflexes from nervous pathways to each muscle via spinal interneurons regulating the ongoing activity and mechanical output of multiple muscles, and longer-latency pathways to spinal networks and higher CNS areas (cerebellum, basal ganglia, and cortex). Both processes are important to estimate state and update internal models to coordinate balance and plan movement (Wolpert et al., 1995; Wolpert and Ghahramani, 2000; Proske and Gandevia, 2012).

emerges in development as a dynamic and spontaneous process through recurrent perception-action loops where knowledge of the external world is integrated with knowledge of self-movement (continuous exploration of the infant's own body) as the body moves through a force field.

Findings from behavioral brain research also provide evidence for this perception-action coupling. Alaerts et al. (2007), by means of a tracking task, show that proprioception is subject to constraints from extrinsic and intrinsic reference frames that are continuously updated[2].

Building upon these theoretical and empirical perspectives, we propose that PK is organizationally integrated as a coupled system, not restricted to the constant activation from deformations of the dynamic body to produce sensations about the position or the movements of the limbs (Sherrington, 1907; Fourneret and Jeannerod, 1998; Hewett et al., 2002). Thus, the central nervous system would not be unique in its capacity to control the wide variety of action-oriented abilities. Rather, these abilities would arise from a systemic regulation, including cortical and subcortical networks, effector organs, sensed environmental constraints, such as gravity and friction (Goodwin et al., 1972; Gapenne, 2014), as well as sensed social constraints, such as those related to social interaction. However, this organization in action remains ambiguous.

PROPRIOCEPTION IN ACTION AS A PUZZLE: IS AN INTERNAL MODEL THE MISSING PIECE?

Most accounts in which proprioception seems to be coupled with kinesthesia, although not explicitly, aim to capture how afferent information is used by the internal brain processes to regulate motor control and coordination. This could be due in part to the fact that it is generally accepted that proprioception in the absence of muscle contraction (passive proprioception) is dependent only on the processing of peripheral inputs (Craggs et al., 1979; Nakajima et al., 2006). Indeed, the relative contribution of well-recognized processes to proprioception when the agent is in action, with muscle contraction with afferent and referent signals (active proprioception), remains unclear (Proske and Gandevia, 2012).

A closer look reveals the striking difficulty that we address in this section: the role of afferent information within the context of movement control and coordination. Theorists supporting internal models for motor control have expressed a clear position in this debate[3]. This is based on a recognition of proprioception as the means to provide the agent with a variety of crucial information for motor learning to occur.

These theories have been used to understand how the agent perceives the difference between self-initiated voluntary own actions (sensory reafference) and passive, involuntary, and unexpected (so-called sensory exafference) movements (Proske and Gandevia, 2012). Voluntary and accurate motor performance depends on self-generated reflexes, from nervous pathways to each muscle via spinal interneurons, and on a predictive CNS internal model to overcome noise in proprioceptive receptor signaling (Wolpert et al., 1995; Wolpert and Ghahramani, 2000). In turn, this anticipatory signal is subtracted from the incoming sensory signal to cancel the self-generated portion (a reafference), and create a neural representation of the outside world (an exafference) (Crapse and Sommer, 2008). Learning occurs as a result of the continued interaction of proprioceptive feedback and motor performance, thus, strengthening the reference mechanism and allowing the newly acquired skill to become part of the agent's repertoire of learned movements. Once a motor skill becomes automatic, its performance is under the control of a motor program. More recent research has generalized this idea by sustaining that an internal prediction of the sensory consequences of our actions—a copy of the motor commands to muscles as a centrally represented movement pattern stored in memory—is compared with actual sensory afference (Mitsuo et al., 2003; Wolpert et al., 2011).

In short, neural control centers are thought to predict and specify the motor commands required for active (self-initiated) movement (Farrer et al., 2003; Capaday et al., 2013). These rich internal models work similarly to a B-format; they "represent states of the subject's own body and, indeed, represent them from an internal perspective" (Goldman, 2012, p. 73). Briefly, they are doing all the functional work of proprioception regardless of the role of the body and its relationships.

We, however, believe that this may be problematic. The motor command specifies a precise value for a parameter of position, speed, or other, a corresponding unique value at the level of the sensor, with the variations being totally determined (Piaget, 1937; Lenay, 2006). As Gapenne (2014) asserts, this hypothesis emphasizes the existence of a bijective relation between action and sensation in the case of proprioception that "primes the subsequent inferences realized by the 'brain,' [which] are produced 'at random' remains mysterious [. . .] Where do these commands come from? Why do they take the form that they do? Are they generated by a 'program'?" (Gapenne, 2014).

In contrast to this position, we could think that active proprioception—in a PK system—is something the agent does in a particular situation and in an ongoing fashion. For example, it is certainly relevant in the motor control for an active human agent to walk on a swaying tightrope or for a spider caught on her windblown spiderweb. Both must fine-tune their muscle activity to maintain posture, coordinate sequential movements involving multiple joints, or be prepared for the next move and to stay upright. This motor command would be more than just a matter of pure effectuation that depends on an updated internal representation of body position during the production

[2]Furthermore, neurophysiological research has shown that brain area activations during passive and active proprioception, as somatosensory evoked potentials (SEP), seem to be both afferent (due to the activation of peripheral afferents) and efferent (the influences of descending pyramidal and extrapyramidal influences; Coquery et al., 1972; Beets et al., 2012).

[3]The internal models for motor control theories—neural mechanisms that can mimic the afferent/efferent characteristics or their inverses—assume that the central nervous system (CNS) is able to prepare in advance to differentiate between these two classes of sensory afferences, by sending a parallel "efference copy" of its motor command to sensory areas (von Holst E and Mittelstaedt, 1950; Kawato, 1999).

of learned movement. In this case, the agent would not be able to have access to any variations other than those produced by their own actions—an idea that denies the importance of the various forms of activity of the sensor interacting with a dynamical environment.

From this point of view, phenomena such as gravitation or friction always leave a certain degree of uncertainty concerning the movement which will actually occur (Henri, 1902). These variations, as Gapenne (2014) claims, which cannot be determined by the command, are actually a condition for the possibility of constituting an experience of the spatiality and temporality of the body/self in the present time, or toward accurate coordination with the environment on the basis of the constant and actual variations. This is true even when, as we have already stated, this PK perceptual experience involves the full set of sensory organs.

There is some evidence in support of an interpretation of PK-coupling in sensorimotor theory terms. For instance, a study in which subjects were asked to apposite the index fingertip of one hand to that of the other hand, found that the index fingertip was localized with equal accuracy and with no greater variability when the hand was moved actively by the subject or passively by an experimenter (Darling et al., 2018). The study found the differential activity of the sensor when interacting and no evidence that accurate proprioceptive localization or motor performance depended on the predictions of a CNS internal model to overcome noise in proprioceptive receptor signaling (Darling et al., 2018).

Consistent with this finding, studies conducted in light of the theory of referent control of action and perception (Asatryan and Feldman, 1965; Feldman, 2016), propose that to produce intentional motor actions, the nervous system changes specific neurophysiological parameters—the spatial thresholds at which muscles begin to be activated. When changed, these parameters shift the equilibrium state in the interaction between the organism and the environment[4]. Therefore, these parameters do not result only from the meaningful perception of the B-format, but also from the perception of proprioceptive-kinesthetic coupling with the body situated in the actual environment, with dynamic possibilities for action and oriented with respect to the direction of gravity. As Feldman (2016) proposes, the emergence of optimal sensorimotor action happens without preprogramming due to the cooperative tendency of neuromuscular elements to reach the shifted equilibrium state.

Based on this type of evidence, and moving forward to internal model descriptions, we argue that proprioception goes beyond a positional sense and the preprogramming of motor commands. The PK system would be the origin of spatial frames of reference in which neuromuscular elements are commanded to work (Feldman, 2016). Moreover, in the distinction between active and passive movement, we assert that the agent, with her own activity, is sensitive to the effects of her own actions and to the variations of the afferent signals. This moto-proprioceptive coupling allows the emergence of a continuous and dynamic reference to calibrate other sensorial signals through action (Iscla and Blount, 2012; Lebois et al., 2012). Accordingly, Gapenne (2014) supports that the singularity of proprioception lies in the fact that it is a firm reference-point, a mechanism of "filtering and calibration," which allows an agent to dissociate between self and world, by attributing variations either to her own activity (and thus to the effects of her actions) or to events over which she has no control (Henri, 1902; Gapenne, 2010).

ESMT provides us with a more coherent account of these conceptual issues and findings, taking into consideration agents acting in everyday life, crossing their arms or walking fast to get to work, or avoiding losing their balance when the subway makes a sudden stop. The agent continuously tries to adapt to the disturbances and to recognize meaningful interactions. Noticing this PK coupling nature in perceptual experience and developing a framework unconstrained by the limitations of the current accounts, will be the goal of the rest of the paper. In the following sections, we propose how a description based on ESMT, with deeper links to phenomenology, can contribute to a better understanding of the PK perceptual experience in active agents.

OVERCOMING THE BIAS: THREE KINDS OF PROPRIOCEPTIVE-KINESTHETIC CONTINGENCIES (PK-SMC)

In a similar way to the ecological approach, in the enactive approach to cognitive science "perception does not consist of the recovery of a pre-given world but exists rather in the perceptual guidance of action in a world that is inseparable from our sensorimotor capacities" (Varela et al., 1991, p. 17). This view rejects mainstream theories of perception, which claim that perceiving is about giving rise to internal mental representations from the external world. In this respect, Varela et al. (1991) realizes that a foundational concern in developing this theory, which replaces representations with embodied action, is "to determine the common principles or lawful linkages between sensory and motor systems" (Varela et al., 1991, p. 173). Indeed, cognition is understood as a hands-on practical activity taking place in concrete situations (Varela et al., 1991).

ESMT, as a philosophical and scientific research program (e.g., O'Regan and Noë, 2001; Noë, 2004; O'Regan, 2011), has been developed with a similar concern[5]. Accordingly, perceiving is a bodily skill exercising an implicit know-how

[4]Feldman (2011) proposes the equilibrium-point theory in response to the posture-movement problem: why active movements away from a stable posture are not opposed by stabilizing mechanisms, rather than being specific neural structures representing spatial frames of reference selected by the brain.

[5]This theory is often considered part of 4E Cognition perspectives in cognitive science, where cognition is embodied, enactive, ecological, and extended. However, the heterogeneity of the different lines of research within 4EC has led to certain disagreements that have partially split their bond. It is not the purpose of this article to delve into those disagreements, and for our purposes, we will consider that they share these basic claims, and, therefore, we can raise them as bonded (Gonzalez-Grandón and Froese, 2018): (a) cognition depends on the characteristics of the agent's body and its interaction with the physical and social environment, and (b) the rejection of mainstream theories, which treats proprioception, for instance, as a passive process of sensorimotor information processing in the brain to set up detailed internal mental representations of the body and its parts (Bermúdez, 2000; Gallagher, 2003; Cardinali et al., 2009).

of the systematic ways that sensations change as a result of potential movements, that is, of sensorimotor contingencies (SMCs) (O'Regan and Noë, 2001; Silverman, 2018). Thus, perceptual modalities differ because they relate to a particular set of exploratory bodily movements: visuo-motor, auditory-motor, proprioceptive-kinesthetic, etc., which together constitute a detailed, directed, and unmediated awareness and allow access to the environment. Stemming clearly from a background of ecological laws, the properties of the SMCs related to the environment are the most general kind of regularities or so-called "laws" of SMCs.

In the following, we suggest that a felt PK perceptual experience is inseparable from sensorimotor expectations. We describe these PK-contingencies as depending on the awareness of the self's potential actions and interactions, abilities that an agent may acquire over a particular history of learning within a specific ecological and self-other environment.

Phenomenology and PK-SMCs

As a means of distinction, ESMT is not only an account of the lawful linkages between sensory and motor systems involved in perception; it also has set itself the much more challenging task of explaining the felt aspect of phenomenal consciousness. It assumes that experience is not caused only by some internal correlate, such as a B representation; in the words of Myin and O'Regan (2002, p. 33): "phenomenality is not caused by some brain process, but is constituted by the different capacities that 'feeling' involves."

But what is special about the proprioceptive and kinesthetic conscious experience that makes it different from other mental phenomena, such as inference thought or color perception? To some extent, when framing the phenomenology of bodily awareness, we can consider the difference between not paying specific attention to our body and actually feeling an exasperating itch in the right leg. In this respect, proprioceptive awareness has been found in philosophical literature related to three domains of experience: the sensation of body position or the sensation of the location at which I feel my hand making the sign of peace occurring (sensorial information from specialized mechanoreceptors); first-person experiences of the sense of body ownership (the awareness of the hand that making the sign as being my own); and ecological self-experience, which is described as the ability to converge many relational aspects into a coherent identity (De Vignemont, 2018).

In particular, in this section, we are motivated by the domain of experience about what is it like to feel one's limbs along with their possibilities for action as one's own? So, we make a critical remark on the view that the felt location of bodily sensations suffices for the sense of bodily ownership (Crane et al., 1992); we favor the possibility that the phenomenology of ownership is over and above bodily sensations and that it is rather a feeling of bodily presence, as De Vignemont (2018, p. 44) proposes: "For instance, when something brushes our knee, not only do we feel a tactile sensation, we also become suddenly aware of the presence of our knee as being located in egocentric space, as a body part that we can reach and grasp. The existence of such a feeling is well-illustrated by amputees who still feel as if their lost limb were still there, physically present."

This proposal is close to holding an action-based theory of perception, as an ESMT view of perceptual awareness. Indeed, the notion of the feeling of presence has originally been proposed from ESMT to characterize the detailed visual phenomenology associated with actual integrated scenes, even though the depicted scenes are not co-present at once (Noë, 2004). Feeling a body as present involves being aware of it as a whole object located in space and time, such as a sponge that one can explore from different perspectives and that one can actively manipulate. It is true that ESMT is particularly compelling for the visual and auditory modalities; however, the inherently exploratory nature of PK perceptual experience helps to account for the fact that PK perceptual experiences have a special phenomenal quality, that is not shared by other mental phenomena; and we can clearly see how perception-action coupling enriches the perceptual experience.

Thinking in PK perceptual experience as a feeling of bodily presence may provide powerful reasons for thinking that PK perceptual experience is constituted as the exercise of an exploratory bodily skill, which is refined as a result of expertise. Whenever the agent is effecting an actual change by self-movement, it has the effect of improving the veracity of attentive and sensible perceptual experience by confirming the anticipated sensorimotor regularities. Furthermore, if the PK conscious experience is constituted by potential exploratory movements it may turn out to be misleading, which has been amply demonstrated in the case of the bodily illusions when being wrong about own body's sensations and body awareness, such as in the Pinocchio illusion (Lackner, 1988) and rubber hand illusion (Botvinick and Cohen, 1998).

This solid connection between perceptual experience content and possibility for action is not new; it is crucial in Merleau-Ponty's "Phenomenology of perception" (Merleau-Ponty, 1945), in Gibson's affordance conceptualization (Gibson, 1977), and in Dreyfus's description of perception as a skill (Dreyfus, 1996). Here is where skill theories provide a route to naturalizing phenomenology: in this view, perceptual experience is not caused only by internal models but consists of various abilities that organisms have to feel, sense, move, grasp, respire, and interact. In order to explain the experience, therefore, instead of searching for neural correlates that ingrain phenomenality into electro-chemical mechanisms within the central nervous system, it is necessary to describe each of the different abilities that the organism displays when it engages in the perceptual activity.

Perceptual experience is shaped by that ongoing interaction with an environment at a present time, where manifold sensorimotor contingencies are at play. However, clearly not all of that SMCs are accessible to the organism's perceptual awareness at the conscious moment of "now",—Varela (1999) shows that this moment has a duration of 1–3 s. indeed, some of these are realized by associated exploratory movements, and others are left out. As Myin (2016) argues, an organism has acquired, on the basis of a history of interactions, a sensitivity in its perception and action for each interactive generality that consists of implicit know-how.

However, it is not yet entirely clear what this phenomenal basis of PK perception means for the agent's experience. There are at least two possibilities, which we will refer to as perceptual sensitivity and perceptual awareness following Noë and O'Regan (2000) and Noë (2002)'s general distinctions, respectively:

1. PK-perceptual sensitivity: In general, this possibility comes from the habitual perceptual coupling of an organism and environment that lies in the history of previous interactions, that is, in the organism's coupling history with its physical and social world. O'Regan and Noe (2002) identify the sensation with a pattern of skillful activity. In ESMT terms, this means the perceptual experience of mastering sensorimotor contingencies (Froese and González-Grandón, 2019). When referring specifically to PK as a way of doing things, this sensed experience is a basic perceptual sensitivity of knowing how it feels to move the body even if the agent cannot directly sense all their body segments or lengths of joints simultaneously. Following Husserl's "habitual consciousness" conceptualization, Sheets-Johnstone (2019) describes this kind of sensitivity as an ongoing presence constituted by mindful bodies sensing themselves and their habitual relationship to the world.
2. PK-perceptual awareness: This possibility focuses on what the coupling affords, to be aware of each detail, and, although it is the result of the mastery of the relevant SMCs (Noë, 2002). It also consists of being aware of our immediate perceptual access (O'Regan and Noe, 2002). A feeling experience has qualitative dynamics of some individual kind, such as abrupt, slow, unexpected, or contractive, or combined when action or interaction unfolds. Living humans are not consciously aware of everything that their bodies do. But sometimes, when being alerted by something significant, such as a sudden cramp or tremor in one leg, this particular felt quality invites us to choose a particular pattern from among others, allowing it to play a prominent role in the embodied organism's present occurring actions (Myin and O'Regan, 2002; Myin, 2016).

PK-perceptual sensitivity as a possibility implies that specific ways of perceiving involve specific movements. When a person bends over to button up their shoelaces, for instance, she is not aware of each of her precise movements or postures through the ongoing activities. In describing this distinction, Noë (2002, p. 569) makes the following interesting observation: the driver, for example, who fails to pay attention to what he or she is doing or to a that to which he or she is responding to is still able to exercise mastery of the sensorimotor contingencies needed to drive the car. Such a driver is, as it were, on "automatic pilot."

However, the possibility of PK-perceptual awareness is a matter of it being able to deploy a potential skill, namely integrating one's perceptual skills into one's intentional and spatio-temporal present action. This would imply that the agent is currently attending a sensorimotor contingency that has been previously learned. Moreover, following this distinction, the traditionally "intentional access" is not described in subpersonal terms anymore, as is the case with weak approaches. We may think about the possibility of accepting qualitatively different accounts: there must be some corporal mechanisms that are responsive to proprioceptive information from the entire body all at once, but others that differentially select between bodily parts. Then, as Fridland (2011) affirms, it seems that the PK conscious experience would be of multiple objects and would depend on the history, interests or plans of an agent. Although it would be rare to imagine proprioceptively and kinetically attending to the entire body in all its detail at once, following these ideas, it could be achieved with training.

Being more specific, coming from ESMT, PK knowledge-how may be identified with bodily skill rather than with possessing a B-format representation. Following the proposed distinctions, skilled PK-perceptual experience can be understood in terms of two key characteristics of PK-interaction, one habitual and the other more attentive, both presenting some kind of continuity, which is evident in perceptual learning. That is, PK perceptual experience is claimed to be constituted by the bodily skill of knowing how proprioceptive/kinesthetic sensations would change as a result of potential overt body movements. This is where implicit know-how constitutes this experience in terms of the perceptual accessibility of the currently non-accessed detail, and explicit know-how constitutes the highly attentive experience that assesses which potential PK-SMCs we should become aware of.

PK-Phenomenal Experience and Some Pieces of Evidence

Given the issues raised above, if PK awareness is to qualify as a legitimate form of awareness and not just subpersonal information, we can follow O'Shaughnessy (1995) and Fridland (2011) when arguing against having two separate explanations for conscious and subpersonal proprioceptive processing. From a phenomenological and ESMT stance, PK perception is not only about whether there is "something it is like" to experience parts of the body as own, such as a "sense of body ownership"[6] but an immediate and direct first-hand or first-body experience with a felt qualitative dynamics.

Husserl (1989) describes the kinesthetic experience in terms of its qualitative nature: the dynamics of movement. In this sense, Sheets-Johnstone (2020) may reinforce the position in which it is not just a pre-reflective awareness of own body that is not very detailed, as proposed by Gallagher and Zahavi (2012, p. 155): "these postural and positional senses of where and how the body tends to remain in the background of my awareness; they are tacit, recessive. They are what phenomenologists call a 'pre-reflective sense of myself as embodied'." Instead, consider Sheets-Johnstone's description: "When we move, we kinesthetically feel the dynamics of the movement as they unfold, and insuppressible qualitative dynamics. A specific sensuous quality is indeed kinesthetically experienced" (Sheets-Johnstone, 2020).

[6]Some theories of the sense of bodily ownership try to address how an agent goes from a proprioceptive experience with the non-conceptual content to a proprioceptive judgment with the conceptual content (Peacocke, 2014). In this sense, (Gallagher, 2003, p. 3) contrasts the typical and "non-reflective awareness of the body," with proprioceptive awareness as an introspective or reflective type of proprioception. However, we think that we do not need to overly intellectualize human embodied experiences in order to classify them as genuinely perceptual.

In fact, following Husserl and her position, the description of the PK perceptual experience becomes more robust as it comes along with a sense of body posture and movement relative to the interaction with the environment. The agent feels a PK sense of her own body parts and their potential movement in relation to something or someone. In this regard, this view is much closer to the notion of "ecological self" from Neisser (1988) when describing this PK sense of dynamical self as an interactive body to produce sensations about the own movements in the ongoing interaction.

Consider the following basic example: when crossing your arms it is not simply necessary to register where your arms are positioned in space for the sake of knowing where your arms are as if you were solving a problem. Rather, this is a directly perceived and pragmatic problem: if you want to give someone a hug, you have to know what position your arms are in, how far or close the person you want to hug is, how much friction you have in terms of the clothes you are wearing, if the ground you are standing on is tilted, etc. This does not involve a theoretical reflection but a characteristic PK perceptual know-how: your bodily action is ready to go. PK accounts for one's ability to detect limb position and bodily posture from the inside, and it consequently has to be in a constant relationship with ecological interaction.

In a nutshell, this strong embodiment thesis helps us to describe in greater depth what PK-coupling feels like; it considers that this experience is about a spatio-temporal presence and is foundationally grounded in the skilled kinesthetic body (Sheets-Johnstone, 2020)[7].

We can already note that these theoretical possibilities, in the framework of PK on the neurophysiology of motor behavior, attest to the importance of body awareness in proprioceptive perceptual learning. Feldman (2016), when referring to self-initiated movements at which muscles begin to be activated, rather than giving an absolute role to the afferent feedback, suggests that the central influences on the neuromuscular periphery (motoneurons) have an interactional and dynamic dimension.

There is also evidence, considering the unloading reflex—the reflex inhibition of the muscles of mastication that occurs when food or other material between the jaws suddenly collapses and helps to stop the jaws forcefully coming together—as an example of involuntary action. Ilmane et al. (2013) demonstrated that the corticospinal and other descending systems maintain the referent position of the wrist during unloading, thus, allowing the neuromuscular periphery (in the continuous and dynamic organization with central influences) to change motor commands and the wrist position in response to unloading, as an external and surprising perturbation.

[7]Our PK phenomenology proposal has a lot of assumptions that are by no means universally accepted. We limit ourselves to highlight that we ought to consider alternative methods for understanding and distinguishing the nature of such PK experience and their relation to our proposal. On our side, we think we ought to rely on both first and third-person perspectives—phenomenological descriptions of experience contrasted with naturalistic explanations—in order to come up with accurate and useful categorizations of these conscious states. Such a dialogue platform will be more likely to yield a solid theoretical PK experience foundation.

Another source of evidence that is consistent with these findings comes from the kinesthetic illusions elicited by the tonic vibration of the tendon of an elbow flexor (Eklund, 1972; Goodwin et al., 1972). Vibration enhances the activity of flexor spindle afferents, eliciting an illusion of elbow extension as if elbow flexors were stretched. Most interpretations of this illusion argue that it results from an increase in the afferent component, while the central component remains unaffected by vibration. Here, again highlighting the importance of the whole percepto-motor system, Feldman (2016) suggests that the illusion can be explained by the influence of vibration on the central component, resulting in an actual motion-learned and reliable (meta)stable pattern in the sensorimotor coordination (Buhrmann et al., 2013).

Thus, to account for the constitution of this particular felt bodily experience—the immediately felt qualities of the experience of spatial and temporal self-orientation in action, such as in feeling oneself being the one acting, for example—the agent must learn to qualitatively distinguish between three sources of variation in the PK sensory signals that become coupled within an open-loop fashion in the online interaction: PK-SMCs self, PK-SMCs self-ecological, and PK-SMCs self-other.

In the following section we introduce and describe each of these PK-SMCs, analyzing the main conceptual points related to ESMT and kinesthetic phenomenology, and we also offer a formal description of each of them that leads to the development of our PK minimal model.

Proprioceptive-Kinesthetic Sensorimotor Contingencies-Self: PK-SMCs-Self

A key characteristic of a PK system is its sensibility or awareness of its own musculoskeletal parts in relation to other parts of one's body and of their possibilities for action and interaction. The PK-SMCs-self contingencies are described in this regard as involving the exercise of a bodily skill, the know-how of the systematic ways that a sense of the bodily self changes as a result of the potential moving self, in relation to one's body. We propose that all the aspects of the phenomenology of the sense of proprioceptive and kinesthetic coupling are related to both this inherent self-oriented sense in space and in the present time, and also in relation to perception and action cycles in interactions that together comprise the PK-SMCs-self kit. For instance, the experience of sensing the positions of body segments and their possibilities for movement in relation to each other. Certainly, the relational features always involve the physical and social world in the first place, and they do not require internal comparison between B-formats; in this section, however, we will only focus on the contingencies of the spatial and temporal orientation of the body's own parts and its possibilities for action, leaving for the following sections the establishment of meaningful relations between ourselves and the ecological and social world.

Moreover, in addition to the afferent signals of limb position that provide the central nervous system (CNS) with information

about the spatial orientation of the body's own parts, the PK-SMCs-self also involves efferent signals, environmentally sensed constraints, such as gravity and friction, or the sensation of movement of another agent. The PK-SMCs-self are thus constitutive of the sensorimotor exploratory behavior of any human agent, as a form of baseline behavior to the ecological self, and are also enablers of self-other interaction.

The importance of the PK-SMCs-self as felt is also evident in the case of deafferented agents who lack PK perceptual awareness in a large part of their body. Although rare, some viral infections can cause autoimmune reactions that selectively attack the peripheral nervous system and destroy afferent pathways that are part of the PK system (Connell et al., 2008). In these cases, subjects no longer have proprioceptive awareness in the parts of their bodies affected by neuropathy. They lose the ability to immediately recognize their practical possibilities for action. But since this condition does not affect the efferent nerves, and it is still possible for subjects to regain the ability to produce movement with those parts that they can no longer feel but can visually perceive. This had been taken to show that proprioceptive awareness is not necessary for bodily action (Bermúdez, 2000; O'shaughnessy, 2008).

However, we argue that in the absence of the PK-SMCs-self set, ordinary action as we know it is impossible. Deafferented agents have severe problems in the online control of action, and their actions may seem performed distant because of lacking PK perceptual sensitivity and awareness. When a deafferented agent does not sense or feel their limbs and uses her attentive gaze instead, she loses the possibility of experiencing her orientation in relation to the limits of her own body and directly perceiving the possibilities for acting and interacting with her surroundings (Howe, 2018). Certainly, a deafferented subject with a lot of training will be able to achieve better possibilities for acting and interacting, and a form of awareness may arise, but it is not a PK perceptual awareness.

We argue that to recognize the difference between a skillful PK perception, from one that is not, or between the sensitive or aware qualitative dynamics variety, between habitual experience from paying attention to one's muscles movement and interaction possibilities, is a challenge that can be better understood regarding skilled PK-SMCs-self, where one of the two following possibilities must be at play:

- Skilled PK-SMCs-self (SPK): this possibility comes from taking into account the mastering of PK-SMCs-self. A PK-SMCs-self skilled agent has a learned perceptual sensibility, a widely recognized repertoire of body orientation, and concrete action possibilities in particular contexts from which a specific contingency can be selected for attention. This skilled agent therefore also has a PK perceptual awareness.
- Non-skilled PK-SMCs-self (NSPK): In contrast, this possibility comes from considering agents such as those who are deafferented or live with some similar affectation. The PK-SMCs-self have not been developed properly, and the agent thus does not recognize the limits of their own body and the possibilities for acting and interacting with their surroundings in a practical way. As a deafferented PK agent whose perceptual experience is disconnected from their practical possibilities.

One way of shaping these intuitions is to formalize the PK-SMCs-self of an agent with the environment through a dynamic systems approach. There have been a few attempts to define SMCs on a strictly formal basis, although with less emphasis on proprioception. Philipona et al. (2003), for example, trying to deduce the dimensionality of the external space of interaction of an agent, proposed an algorithm to capture the position based on inputs and outputs.

For our purposes, inspired by the work of Buhrmann et al. (2013), we chose some variables to describe the PK coupling, and we made use of a minimal dynamical model to describe the different kinds of sources of variation, the PK-SMCs.

PK-SMC-Self/Model Description

Inspired on the basic model for collective movement proposed by Vicsek et al. (1995), we considered the simplest case of only one agent moving continuously inside a $2d$ square region of length L with periodic boundaries. The agent has developed PK-SMCs, denoted by p. The model assumes that the agent has a constant PK perceptual skill during the dynamics, and p thus does not depend on time.

In general, such a system could be described by the next set of equations regarding the agent's position $\mathbf{x}$ updates according to the following:

$$\mathbf{x}(t+1) = \left[\mathbf{x}(t) + \frac{\xi_1(t)}{p}\right] + \kappa\theta(t+1) \tag{1}$$

The first part of the right-hand side of the above equation shows that the agent, in order to move, must perceive its position in the world. This perception is portrayed by the whole first big parentheses of Equation (1), and it is influenced by three things: the real agent's position $\mathbf{x}(t)$, the agent's PK ability p, and other factors that are not explicitly described in the equation but are implicit in the variable $\xi_1(t)$. These could include both external stimuli and internal mechanisms that do not depend on the PK ability but could modify the agent's perception. Going back to the example of the earring, this variable $\xi_1(t)$ could be an unexpected disturbance such as an involuntary handshake or a shove from another person that could alter the agent's perception of their orientation and could have an impact on the final task of putting the earring into. This variable $\xi_1(t)$ is a random variable taken uniformly in $[-\xi, \xi]$[8,9,10]. Then, if the parameter $\xi > 0$ is low, the perception of the agent depends mostly on its PK ability: if the agent has a good PK ability (high p), their perception of their position would be very accurate, but if they have a poor PK ability (low p), her perception would be wrong; if ξ takes medium values, then the agent's PK ability, if good, could absorb its effect. But

[8]Since we do not have enough prior information about the behavior of this external and internal stimulus that could modify the agent's perception, the adequate distribution to portray them is a uniform one.

[9]The variables ξ_1 are sampled at every time because of these "other factors" influence in the agent's movement at each time step of the dynamics.

[10]The uniform interval takes negative values only because the $2d$ square environment has negative coordinates.

if the agent's PK ability is bad, then ξ could amplify an already bad perception; if ξ is high enough, it does not matter if the agent has a good or bad PK ability, as the effect of ξ will cause its perception to be wrong. Below we will specify what we mean exactly by "small," "medium," and "high enough."

The second part of the right side of Equation (1) updates the agent's direction and, consequently, updates its position. It portrays the fact that the agent also needs to move in order to perceive, as was proposed by Gibson (1977). The agent's direction is given by θ; an angle between $-\pi$ and π, and is defined as:

$$\theta(t+1) = \theta(t) + \frac{\xi_2(t)}{p} \tag{2}$$

In order to sum this angle to the agent's positions, it is transformed in a $2d$ vector defined as $[\cos(\theta), \sin(\theta)]$. The random variable $\xi_2(t)$ is interpreted as before: a random variable taken uniformly within the interval $[-\xi, \xi]$[11]. A Skilled PK-SMCs-self (SPK) then implies that the agent is more aware of their possibilities for movement, and a Non-Skilled PK-SMCs-self (NSPK) implies the opposite. For simplicity, we assume that the length step between updates is given by the factor κ. This ensures that the agent's movement is at a constant velocity in direction of θ.

The minimal model thus incorporates our previous proposal that proprioception is coupled with kinesthesia: the agent senses its body and performs it. Based on this, we predicted that an agent with SPK will be better aware of this own position in space and movement possibilities; as a consequence, its future movement will be less erratic than an agent with NSPK.

In order to illustrate the last affirmation, **Figure 1** compares the trajectories in the space of a SPK agent and NSPK agent. As we explain above, the agent's movement will depend on the parameters ξ and p—the combination of which will give us different behaviors. In order to study the effect of each one we first fixed $\xi = 0.5$ and observed how $\mathbf{x}$ and θ changed in time for different values of p. The agent moves in a $2d$ square of length $L = 5$ with periodic boundaries and $\kappa = 0.05$, i.e., it travels 0.05 units in each time step. The total time of the dynamics is $t = 250$. The initial angles and positions to start the dynamics were taken randomly.

Figure 1, top displays the change of θ for different values of p, and we can see that if p is small (=1, blue squares) the agent shows very drastic changes in terms of their angle movements due to the large effects of the external perturbations $[\xi_2(t)]$, implying that the agent does not have the skill to act in harmony with their world. This lack of SPK also influences the agent's spatio-temporal self-orientation; she consequently travels erratically in the space because she does not know her exact position in the world, displaying an erratic trajectory with changes in position and direction (**Figure 1**, bottom Left). This behavior changes as p grows: when $p = 10$ (green filled squares), the changes in θ are not so drastic and the trajectory now shows smaller fluctuations. With these values of ξ and p, the agent is more aware of their spatial position and possibilities for movement, making a somewhat more organized trajectory (**Figure 1**, bottom Center). When $p = 100$ (pink circles), the agent is fully SPK as a result of an active coupling between performing and sensing. The fluctuations in θ are practically nonexistent, and its trajectory is fully organized (**Figure 1**, bottom Right)[12].

Figure 2 shows the change of θ as function of t for different values of ξ and p. When ξ is small (**Figure 2**, top Left), an agent with medium p is SPK, as we discussed above. When ξ increases, high values of p are necessary to reach the SPK. For example, **Figure 2**, top right shows the case $\xi = 2.5$, here an agent with $p = 10$ is not SPK anymore; the changes in its direction are too drastic, it would need a higher p to be a SPK agent. At values of $p = 100$, the agent can resist higher values of ξ; here, the agent is completely SPK and responds well to high values of noise. An analogous situation for this last scenario (of a completely SPK) would be one in which the agent can insert an earring while they are in a moving car on a very irregular pathway or even when their hand is wet and the earring is very tiny.

We can say that this super SPK agent not only has a great PK-perceptual awareness but also high PK-perceptual sensitivity. Her great response to noise and ability to nullify it not only comes from their high PK-perceptual awareness (integrating her purely perceptual skills into intentional and spatiotemporal present actions) but also from their PK-perceptual sensitivity, which gives them the ability to respond efficiently and automatically to high levels of noise that could otherwise affect their conscious actions. Then, the PK-awareness and the PK-sensitivity are correlated in the sense that a high PK-sensitivity gives the agent better PK-awareness and, therefore, a super or complete SPK.

From these results, we can say that an SPK agent is one whose parameter p is high enough to compensate for the effects of noise in the skilled exercise and awareness of the implicit know-how of the lawful ways that sensations change as a result of potential movements. This concept will be extended in further sections but whilst maintaining this general idea. The model is based on established theories of SMC in the sense that it follows some of the descriptions set out in previous sections, although we arbitrarily select parameter values depending on the focus of interest.

PK-SMC-Self-Ecological

Proprioception has been largely described either as a subconscious process, as mentioned previously in relation to B-formats, in that it does not typically require directed awareness or attention or even doubted regarding its perceptual nature (O'Shaughnessy, 1995; Sydney, 1996; Bermúdez, 2000). For us, since we are interested in thinking about proprioception coupled with kinesthesia, as a form of awareness or as a percepto-motor skill that can be developed throughout the life of the organism, we emphasize the interactive co-dependence

[11] In general $\xi_1(t) \neq \xi_2(t)$. This is because the things that could change the agent's perception of their position in the world are not always the same as the things that could change the agent's perception of their possibilities for movement.

[12] The large jumps are due to the fact that the space has periodic boundaries: when one coordinate (x or y) in the agent's positions is too close to the boundary, it appears in the other side of the space.

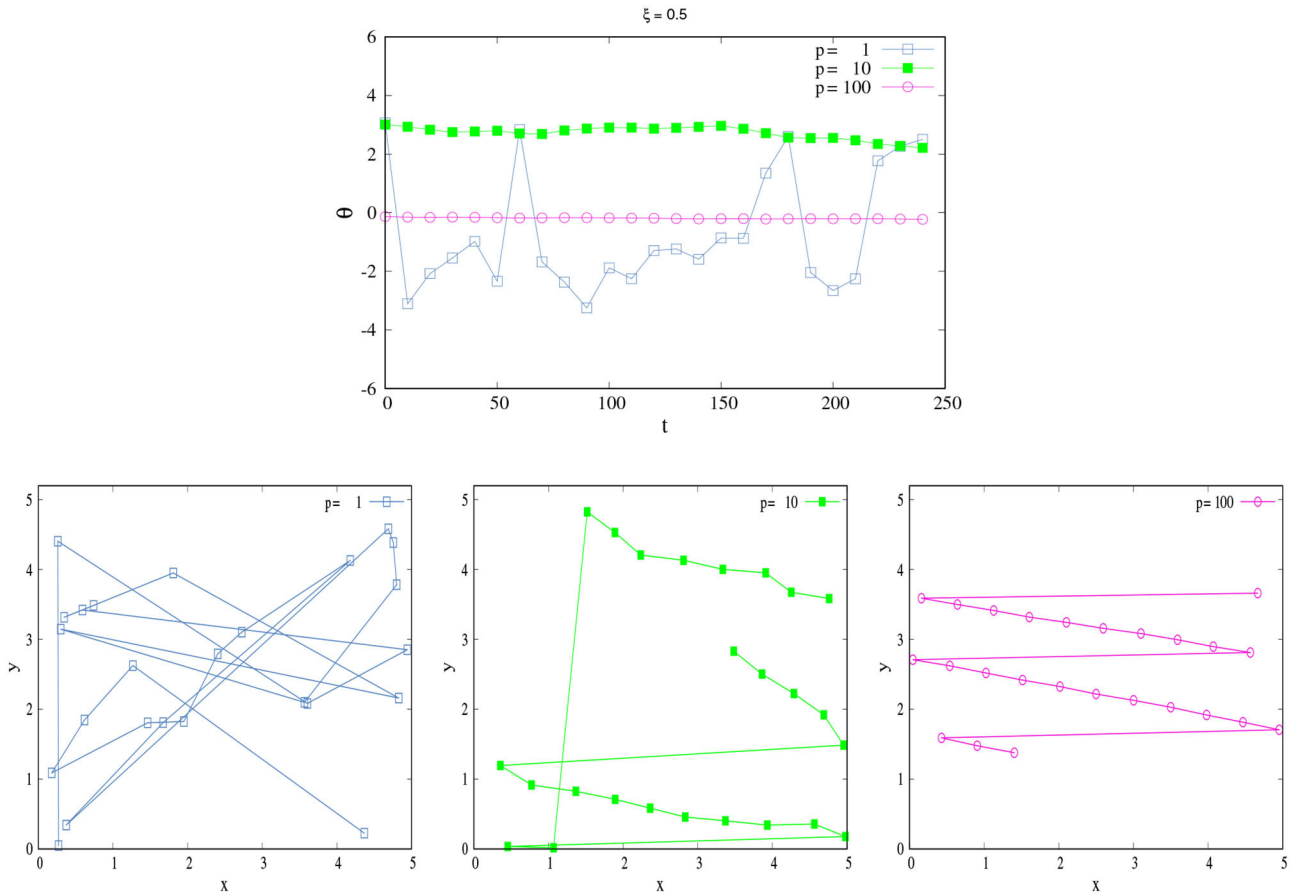

FIGURE 1 | SPK vs NSPK. **(Top)** Change of θ in time for different values of p. **(Bottom left)** Trajectory of an agent with low $p = 1$. **(Bottom center)** Trajectory of an agent with medium $p = 10$. **(Bottom right)** Trajectory of an agent with high $p = 100$. Each trajectory (different color) corresponds to its respective color curve in the changes of θ. Here, $\xi = 0.5$ and $\kappa = 0.05$. All these curves correspond to only one simulation for a single agent in a $2d$ square space of length $L = 5$ with periodic boundaries. Here, $t = 250$, and the curves show every 10 time steps for a better visualization. The large jumps are due to the fact that the space has periodic boundaries; when one coordinate (x or y) in the agent's positions is too close to the boundary, it appears in the other side of the space.

between the PK-SMC-self with the ecological environment that shapes specific modes of coupling. In this line, understanding sensorimotor patterns in a perceptual PK experience becomes relevant for explaining PK awareness as a skill in interaction.

In relation to the distinction made in previous sections between perceptual sensitivity and perceptual awareness Noë and O'Regan (2000) and O'Regan et al. (2004) take this distinction further and propose two other concepts to try to relate these concepts to body sensitivity and body awareness, respectively: "grabbiness or alerting capacity" and "bodiliness or corporality." Similar to the idea of salience in the context of affordance ecological theory, "grabbiness" is associated with the contextual attractiveness of something to a perceiver related to the presence of mastering of SMCs. It also has a complementary aspect, "bodiliness," which refers to how much the perceiver's perceptual awareness will change when the perceiver moves. The greater these changes, the higher the degree of "bodiliness." It is worth mentioning that O'Regan et al. (2005), explicitly state that proprioception does not have "grabbiness":

> "Proprioception is the neural input that signals mechanical displacements of the muscles and joints. Motor commands that give rise to movements necessarily produce proprioceptive input, and proprioception therefore has a high degree of corporality. On the other hand, proprioception has no alerting capacity: changes in body position do not peremptorily cause attentional resources to be diverted to them. We therefore expect that proprioception should not appear to have an experienced sensory quality. Indeed it is true that, though we generally know where our limbs are, this position sense does not have a sensory nature" (O'Regan et al., 2005, p. 60).

First, we consider that the PK system, as a perception-action coupling, does have a sensory nature: the way we position ourselves and move in the world has a particular experienced sensory quality. As Sheets-Johnstone (2019, p. 150) states, action directs attention toward the dynamics of movement that precisely constitute qualitative dynamics, "whether a matter of self-movement or the movement of human and nonhuman animals and of objects in the world." Now, what O'Regan et al. (2005) identified here is certainly the positional component of

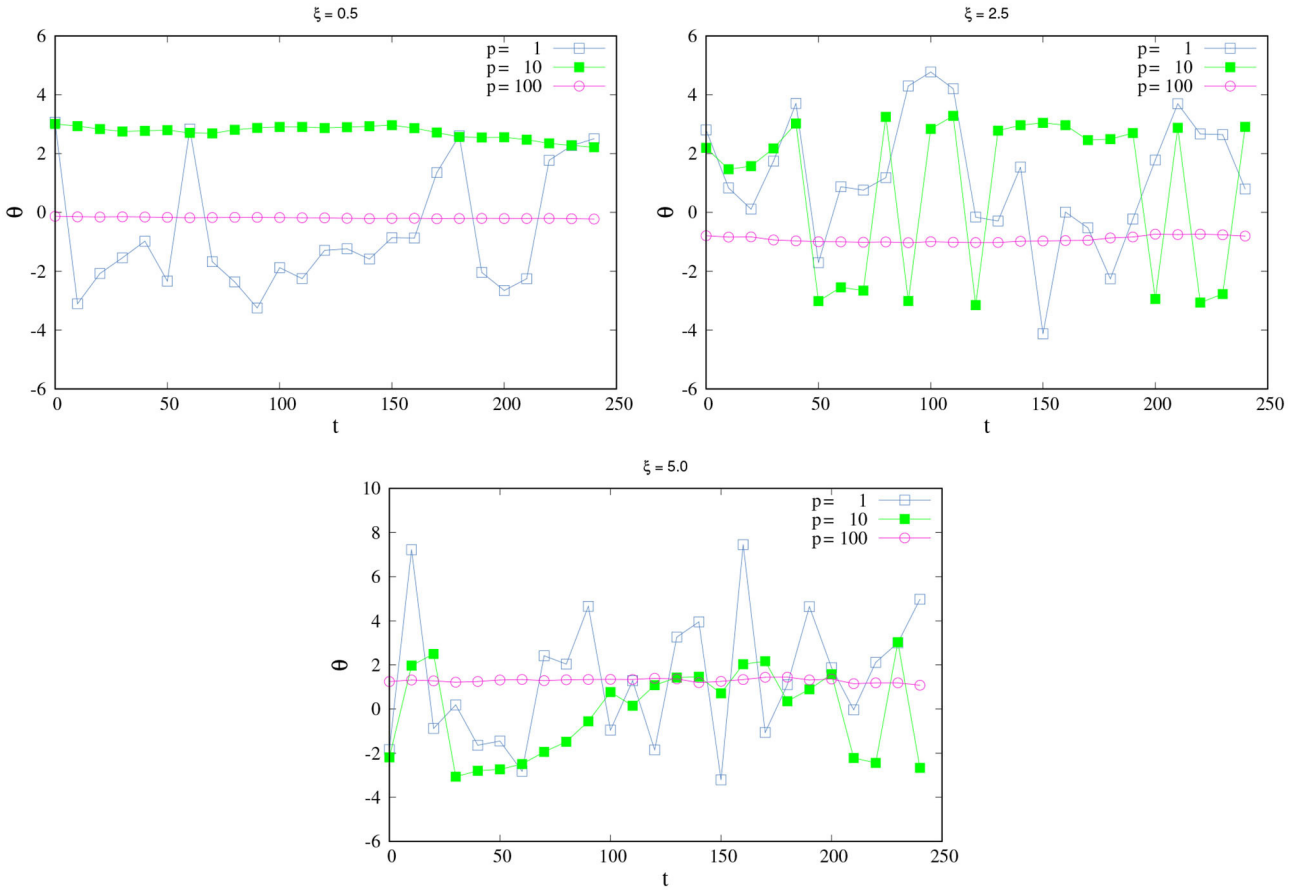

FIGURE 2 | How SPK and NSPK change when noise increases. Change of θ in time for different values of ξ and p. **(Top left)** $\xi = 0.5$. **(Top right)** $\xi = 2.5$. **(Bottom)** $\xi = 5.0$. All remaining parameters take the same values as in **Figure 1**.

the PK system, suppressing the felt or perceived dynamics in the interaction. Whether an infant mastering their PK-SMCs to be able to get into a crawling position on their hands and knees as a form of perceptual sensitivity or body grabbiness or an adult learning a new skill, such as paying attention to a new clinical skill in preparation for medical training, the mastering of PK-SMCs and the acquisition of new skills requires a proprioceptive/kinesthetically-attuned body—a dynamic body that feels[13].

Second, we consider that O'Regan et al. (2005) have left open how are we to understand the relationship between an agent interacting with the environment in a particular scene, such as those where affordances are sensitive to sudden changes in muscular tone or position and activate attentional resources to be automatically directed to the location of change[14]. According to Gibson (1977, p. 140), specific muscles, kinesthetic habits, attentional processes and preparedness, as well as one's own action readiness remain activated throughout the interaction with a particular environment. It is true that it may be less peremptory than in the case of vision or hearing, but grabbiness is also present. Indeed, the claim of ESMT is that the orientation responses primed by the grabbiness of interaction constitute the qualitative feel of PK perceptual experience. In this respect, we argue that PK-SMCs self-ecological also possesses a high enough degree of body sensitivity and awareness with "grabbiness" and "bodiliness."

Drawing on these distinctions, ESMT seems to provide a unique perspective on the consistent description of PK perceptual experience as constituted by a variety of bodily skills. We consider that among human agents, the strategies to be mastered or skilled are always at the interface with the ecological environment and its norms and the social environment.

Indeed, the development or acquisition of particular PK-SMCs describes how an agent becomes attuned to a specific ecological interaction by regulating, selecting (as it is preferable to act more optimally in the known environment), or modulating

[13]Some authors interested in the factors that contribute to the sense of position have reported that position acuity may be improved by increasing the activity of the musculotendinous receptors, for example, by a loaded limb condition (Suprak et al., 2007).

[14]Furthermore, it seems to be following an idea closer to weak embodiment, to B-format notion, where the updating takes place only at an internal level, without requiring an attentive effort in some steps of the learning process.

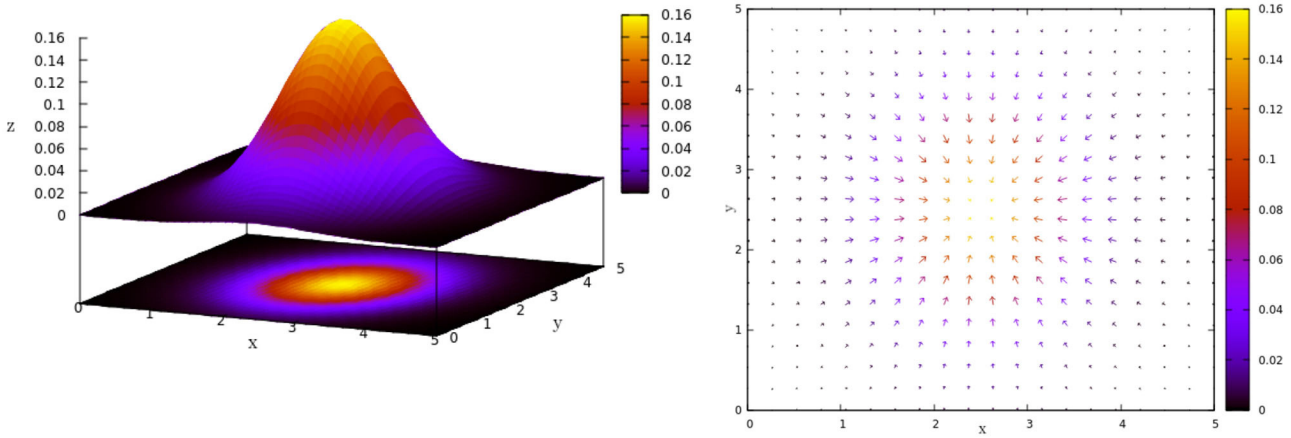

FIGURE 3 | Heterogeneous space. **(Left)** Normal diffusion in $[0, 5] \times [0, 5]$ following Equation (3). **(Right)** Gradient vector field G, associated to Equation (3), which gives us the gradient vector $\mathbf{g}$ at each point of the space.

the relational patterns in accordance with relevant norms. PK-SMCs change as a result of learning and training. That is, it seems clear that proprioceptive awareness is dependent on what we know, how we act, and how we bring attention to our bodies. We refine our feeling of PK-SMCs, providing a pragmatic bodily awareness related primarily to the agent's posture, action possibilities and to constant action and interaction updating as a result of expertise (Gallagher, 2006, 2017; Tsakiris, 2015)[15].

Although our model does not yet include variability in the forms of PK awareness in terms of parameters α and β as functions of p, in future steps of this research, we would like to better understand the qualitative dynamics diversity in the larger differentiation of this ability by including some of these variables in our minimal model.

PK-SMC-Self-Ecological/Model Description

To include the interaction between an agent and the environment in our minimal proposed model, we will consider heterogeneity in space, a concentration gradient that diffuses in a normal way with origin in the center of the space of length L. This implies that for each point (x, y) in the space there is a concentration given by the following:

$$N(x, y) = \frac{1}{2\pi} \exp\left(-\frac{(x - L/2)^2 + (y - L/2)^2}{2}\right) \tag{3}$$

as **Figure 3**, left shows for a space of length $L = 5$.

The agent will interact with this heterogeneous space through each gradient vector in the gradient vector field G given by $G := \{\mathbf{g} = (g_x, g_y) = (\frac{\partial N}{\partial x}, \frac{\partial N}{\partial y}) \ \forall \ (x, y) \in [0, L] \times [0, L]\}$ (**Figure 3**, right). Each gradient vector $\mathbf{g}$ describes in which direction and in what proportion the greatest change in the concentration occurs. To simplify the computations, we consider the normalization of $\mathbf{g}$, i.e., $\mathbf{g} = \mathbf{g}/||\mathbf{g}||$. The new agent's direction $\theta(t + 1)$ will be a weighted sum between the previous direction $(\theta(t))$ and the direction given by the gradient vector $\mathbf{g}$ defined by the agent's actual position $\mathbf{x}(t)$. For this we must modify Equation (2) as follows:

$$\theta(t + 1) = \alpha\left[\theta(t) + \frac{\xi_2(t)}{p}\right] + \beta\left[\theta_{\mathbf{g}} + \frac{\xi_3(t)}{p}\right] \tag{4}$$

with $\theta_{\mathbf{g}} = \arctan(g_y/g_x)$, ξ_3 as a random variable taken uniformly in $[-\xi, \xi]$, and α, β free parameters such that $\alpha + \beta = 1$. Here, the noise variable $\xi_3(t)$ is interpreted as before: an skilled agent will be more aware of the effect of the environment in their movement, following it with more certainty and being able to interact with it effectively. The addition of new parameters α and β portrays the fact that the acting agent may make a distinction between two sources of variation in the sensory signals that affect it: one related to their own activity (α) and another related to their interaction with the environment (β). An SPK then allows the agent to follow (with a certain weight) the direction of the greatest concentration, i.e., the agent has a feeling of a specific type of coordination with opportunities afforded by the various degrees in which she interacts with their environment.

We want to investigate the effect of the PK value p on the interaction between an isolated agent and the environment (PK-SMC-self-ecological). We consider that an agent interacts successfully with their environment if it is capable of finding the origin of the concentration gradient. For this, we suppose that $\alpha = \beta = 0.5$, i.e., the agent takes equally into account in terms of movement, their own direction, and the direction given by the gradient. We are going to consider the average success rate s and the average first-arrival time τ, i.e., how many experiments the agent was able to find the center of the concentration in and how long it took them to do so.

[15]It is not the goal of this article to go into depth in the consideration of many detailed levels of awareness when interacting in different socio-cultural practices. However, it seems certain that learning about particularities of daily life that develop relatively stable patterns of coordination toward a specific practical mode (for the kind of work we do or for games we play) may lead to different levels of PK awareness.

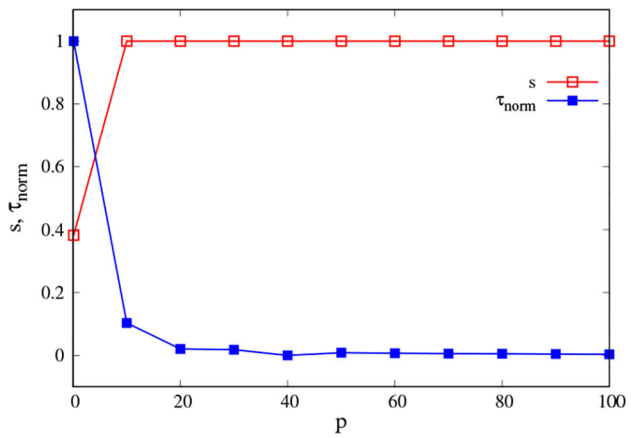

FIGURE 4 | Environmental interaction. Average success rate s (red squares) and average first-arrival time τ (blue squares) as a function of p. The curves are the result of averaging 500 different experiments with $t = 2,500$ each one, with $\xi = 0.5$ and $\kappa = 0.05$. For a better reading of the graph we considered $\tau_{norm} = (\tau - \tau_{min})/(\tau_{max} - \tau_{min})$.

Figure 4 shows the change of s and τ_{norm} as p increases. We see that for low values of PK $p < 10$ the success rate is low (red squares), and the average first-arrival is large (blue squares). This means that an NSA was not always able to find the concentration center; when they did, it took a long time. Their ability to interact with the environment was not good. On the contrary, if the agent has a PK value above 10 (SPK), they are capable of finding the origin of the concentration gradient at every time and also within a very short time in comparison with an non-skilled agent (NSPK). The effect of increasing the noise ξ is the same as before: an SPK agent could become an NSPK if ξ is high enough and their SPK is not sufficient to compensate for its effect in their spatio-temporal self-orientation in present action and interaction. We have explored the effect of α and β in more depth in the next model section.

PK-SMCs Self-Other: Can Sensorimotor Contingencies Account for Processes Such as Social Perception?

The aforementioned idea of ecological PK-SMCs can also be applied to the PK perception of another person. From enactive social cognition, it is known that the motor system is involved in social perception (Gallagher, 2009; Froese et al., 2020). More accurately, in line with ESMT, it has been suggested that social perception consists of the skillful co-regulation of participatory social interaction (De Jaegher et al., 2010). Each person needs to have knowledge of the qualitative dynamics caused by the other's bodily movements concerning their own possible bodily movements. The mastery of these "self-other contingencies," as McGann and De Jaegher (2009) call it, provides a PK-self-other perceptual experience.

According to the strong position defended in this article, both social and ecological PK perception depends on skillful regulation of interaction with different invariants and qualitative dynamics. In each case, this includes perceiving the air as air or another person as another person. However, in this second form, intentional access or perceptual awareness additionally depend on a complementary skillful response by the other person. Both have to master PK-self-other contingencies. If the other agent does not respond appropriately, the PK perceptual experience would be more akin to that of ecological PK perception. Nevertheless, it is not yet entirely clear what this self-other basis of PK perceptual experience means for the agent's experience. There may be many instances for meaningful PK interaction, but we will concentrate mainly on two for the operational purposes of the description and the proposed model. We will refer to these as "PK-self-other sensitivity" and "PK-self-other awareness" forms of PK social perception, respectively:

1. PK-SMCs self-other sensitivity: In this case, one agent's perception of the other agent is only partly constituted by their ongoing social interaction, and each agent's perception can be molded by the other's movements possibilities but without constituting a meaningful shared moment of joint attentive experience. An example includes PK perceptual self-other sensitivity that may be evident in active daily interactions, which often require the agent to recognize the possibilities for the other to act and what their next move will be[16].
2. PK-SMCs self-other awareness: This form gives rise to a jointly attentive unfolding experience because both agents have a mastering of PK-self-other contingencies. The more aware you are of those learned sensitive interactions, the more skilled you are in mastering self-other contingencies. In this case, there is a PK-SMCs-self other perceptual awareness in each agent to realize an attentive, skilled, and participatory performance. For instance, dancers of Argentinan tango can fluidly improvise together only when they actively explore their partner at every moment and reciprocally make their bodies amenable to being sensed (Kimmel, 2013)[17].

What is important in this sensitivity and awareness context is to recognize not simply that during a human's history of coupling, others populate their self-dynamical space action possibilities or act as a reference point for the person's orientation in the present action, but that such interaction may also play a constitutive role in shaping human perception-action cycles and experiences. Indeed, an appropriate PK-self-other experience depends on adequate PK-SMC-self and PK-SMC-self-ecological. We propose that agents engaged in dyadic relations and particularly those having common PK-self-other awareness skills, are more easily able to include other agent's ecological self-action possibilities in their own ecological self.

We investigate these distinctions as a kind of minimal social interaction, arguing that PK self-other contingencies are constitutive of the varieties of PK-self-other experience, either

[16]In this sense, Sheets-Johnstone (2019) propose that from the first social interactions (e.g., newborn-caregiver), the agent incorporates the dynamic flow of body proprioceptive and kinesthetic signals (PK-SMCs of gestures, gaze, gait, etc.) from others into how they modulate their own actions.

[17]"From an enactive viewpoint, other bodies dynamically interpenetrate our own bodily actions and thereby provide a flux of resources for orienting our actions" (Kimmel, 2013, p. 313).

in their sensitive or awareness qualities. That is, we assume that detecting the presence of others is a PK-SMCs-self-other that can be mastered and learned skillfully. Moreover, a skilled PK-self-other contingency is evident in activities like the above-mentioned dance or in sports that require interaction and trained interdependence to ensure a successful outcome. For example, the so-called alley-oop in basketball is an offensive play that requires both teammates involved to sufficiently know and feel the others' moves, one of them throwing the ball near the basket to the other teammate who jumps, catches the pass, and makes a basket (Doeden, 2014).

We advance in our minimal model proposal, based on the idea that an agent performing a jointly attentive unfolding experience directly incorporates ecological information relative to the agents in its ecological self-action possibilities, with PK-SMCs-self other awareness and sensorimotor learning.

PK-SMC-Self-Other/Model Description

The minimal PK model introduces social interaction considering two agents in space. Each agent i has its own PK value p_i and an interaction radius r. This interaction radius portrays the maximum reach of the agent's limbs. The position of agent i ($\mathbf{x}_i$) updates as Equation (1), and its angle θ_i is as follows:

$$\theta_i(t+1) = < \theta_i(t) >_r + \frac{\xi_2(t)}{p_i} \tag{5}$$

where $< \theta(t) >_r$ is the average angle inside of the interaction radius r of agent i (counting itself) and is given by $< \theta(t) >_r = \arctan(< \sin(\theta(t)) >_r / < \cos(\theta(t)) >_r)$.

The role of PK is interpreted in the same way as before: an SPK implies that the agent is more aware of their own orientation and their own activity when interacting with others. The agent has also developed PK-SMCs self-other awareness; an NSPK implies the contrary—that the agent has only developed PK- SMCs self-other sensitivity. The SPK agent will be also, and by consequence of its SPK ability, coordinating its movements with its partner when interacting.

In the case in which we consider the interaction between agents and the interaction of each one of them with the environment, θ_i is updated as follows:

$$\theta_i(t+1) = \alpha \left[< \theta_i(t) >_r + \frac{\xi_2(t)}{p_i} \right] + \beta \left[\theta_{\mathbf{g}} + \frac{\xi_3(t)}{p_i} \right] \tag{6}$$

For the results shown below, we consider the simplest case in which only two agents move inside a square-shaped cell of linear size L with periodic boundary conditions. The agents are characterized by points moving continuously in the plane, and (as we discussed before) they have several capabilities:

- Each agent has an interaction radius r=1 centering in the agent's position $\mathbf{x}$. So, if $d(\mathbf{x}_i, \mathbf{x}_j) \leq 1$, the agents will interact between them, where $d(\mathbf{x}_i, \mathbf{x}_j)$ is the euclidean distance between positions of agent i and agent j, with $\{i, j\} = \{1, 2\}$.
- Each agent i has the ability of PK denoted by p_i. Here, we consider that $p \in [0, 100]$.

Given these minimal assumptions, we remember that agents update their position as follows:

$$\mathbf{x}_i(t+1) = [\mathbf{x}_i(t) + \xi_1(t)/p_i] + \kappa\theta_i(t+1) \tag{7}$$

with

$$\theta_i(t+1) = < \theta_i(t) >_r + \frac{\xi_2(t)}{p_i}$$

in the case of PK-SMC-self-other, and

$$\theta_i(t+1) = \alpha \left[< \theta_i(t) >_r + \frac{\xi_2(t)}{p_i} \right] + \beta \left[\theta_{\mathbf{g}} + \frac{\xi_3(t)}{p_i} \right]$$

in the case of the influence of PK-SMC-self-other and PK-SMC-self-ecological.

In most of our simulations, we will use the simplest initial conditions: (i) at time t = 0, two agents are randomly distributed in space, (ii) they have the same absolute velocity κ, and (iii) they have randomly distributed directions θ. The directions $\{\theta_i\}$ of the agents are determined simultaneously at each time step, and the position of the $i-$th agent is updated according to Equation (7). The value of parameter L (size of movement space) was taken equal to 5 for all shown simulations. For this value of L, the results shown here are valid for $\kappa \in (0.001, 0.1)$, and we used $\kappa = 0.05$ for all graphics shown.

Our first main goal is to find the conditions under which the agents are capable of coordinating their movement (PK-SMC-self other). We measure the success of this simple task by calculating the average velocity v_a proposed in Vicsek et al. (1995) as follows:

$$v_a = \frac{1}{2\kappa} \left|\left| \sum_{i=1}^{2} \mathbf{v}_i \right|\right| \tag{8}$$

with $\mathbf{v}_i$ as the vector defined as $\mathbf{v}_i = \kappa(\cos\theta_i . \sin\theta_i)$ and $||.||$ as the norm function. If $v_a \approx 1.0$, we can say that our agents were capable of performing the task of coordinating successfully; if this is not the case, they failed it.

The upper panel of **Figure 5** shows the change of v_a as a function of ξ for different values of p. Here, we supposed that both agents have the same ability of PK, i.e., $p_1 = p_2$. We can see that values of ξ close to zero, even the lower values of p (= 1), achieved coordination. In another way, for larger values of ξ(> 3), even the agents with high PK p (= 100) are not able to coordinate their movement. Those values of ξ that are of interest are those in which $0.5 \leq \xi \leq 2.5$, as in this range the effect of p is consistent with what we know about PK: individuals with high p (SA) are aware of their position in the world and recognize their possibilities for coordination.

The lower panels of **Figure 5** shows the effect of noise in v_a as a function of (p_1, p_2). The different color maps show the combination of the values of p_i for which the agents are, or are not, coordinated. Here we can see that, for low values of noise (**Figure 5**, bottom left), the only values of p_i that impede a successful task are those that are really low ($p_i \leq 10$). It is enough that one of the agents has this value of PK for coordination not to be reached regardless of whether the other agent has a very good

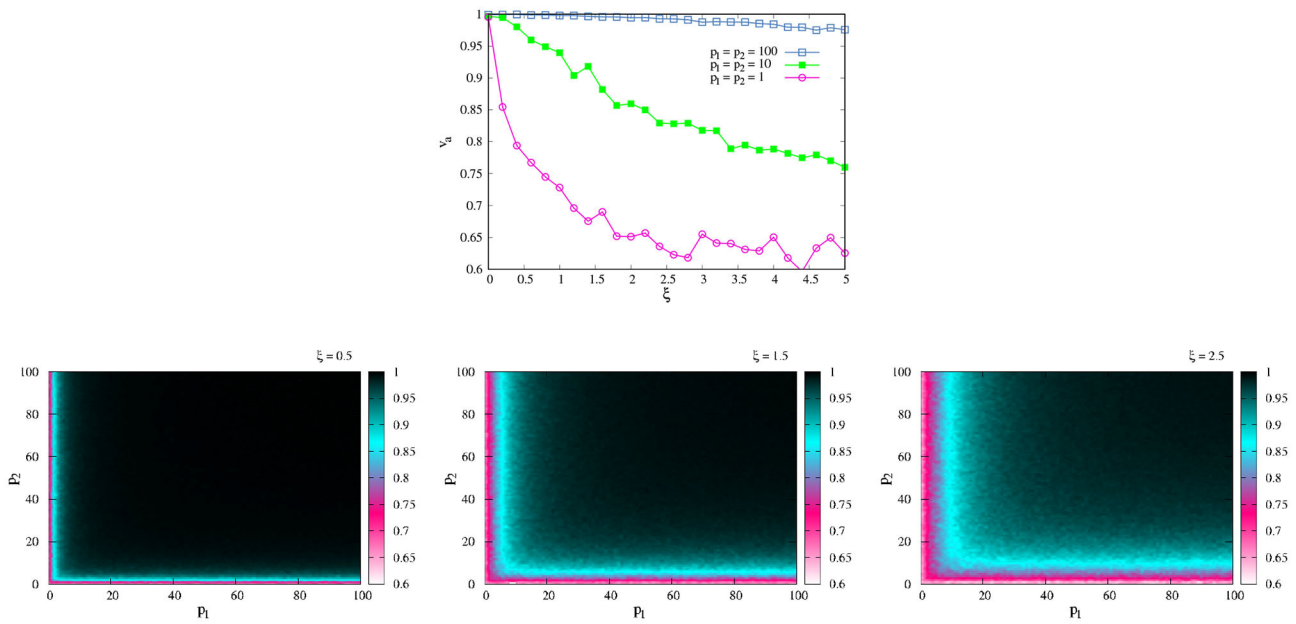

FIGURE 5 | Effect of noise in coordination. **(Top)** v_a as function of ξ for different values of p. Here, we suppose that $p_1 = p_2$. **(Bottom)** v_a as function of (p_1, p_2) for different values of ξ. In all cases the graphs show the average 500 different experiments, each one with $t = 2,500$.

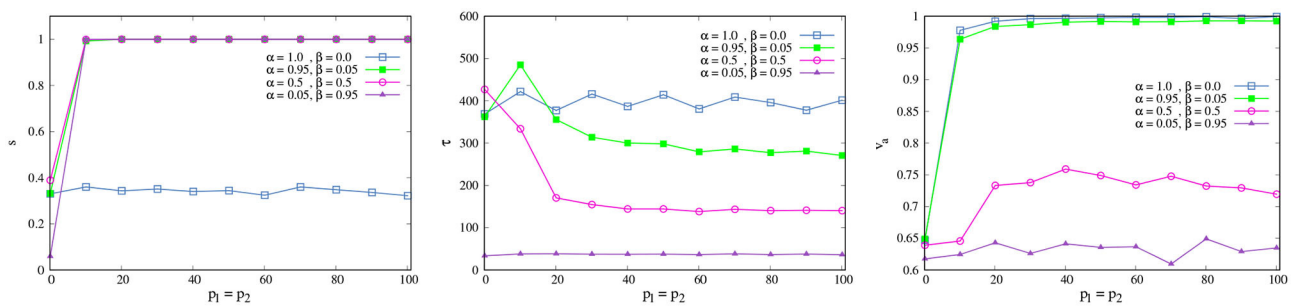

FIGURE 6 | Environmental interaction. Average success rate s **(Left)**, average first-arrival time τ **(Middle)**, and average velocity v_a **(Right)** as functions of p $(p_1 = p_2)$ for different combinations of α and β. In all cases, the graphs are the result of an average of 500 different experiments, each one with $t = 2,500$.

value of p_i (pink and blue zones). On the contrary, if an agent with a PK that is not too low, or medium PK, interacts with an agent with high PK, both end up coordinating their movement (black zone). The effect of noise in decreasing PK values (v.g.r. **Figure 5**, up green curve) then disappears by the interaction with agents with better ability. The left two panels (**Figure 5**, low Center and Right) show similar results for higher values of ξ, and it is clear that if noise increases, the pink and blue zones in the color map are bigger, and larger values of p_i are necessary to achieve coordination. From here we will consider, in the rest of the results, $\xi = 0.5$, which is the value in which the impact of p is clearer.

Finally, we investigate the effect of p, α, and β not only on the ability of an isolated agent to find the center of concentration but on the ability of two agents to successfully interact with their environment and interact between them and to coordinate their movement (PK-SMC-self-ecological and PK-SMC-self-other). The task is to find in a coordinated way the center of concentration.

Figure 6 shows the change of s (Left), τ (Center), and v_a (Right) as functions of p for different combinations of α, β. Here, we supposed that $p_1 = p_2$. We see that when $\alpha = 1$ and $\beta = 0$ (blue squares), the agents are capable of coordinating for $p \geq 20$. Their ability to always find the concentration center ($s \approx 0.4$) is, however, very low, and when they can do it, they take a long time ($\tau \approx 400$). On the contrary, when $\alpha = \beta = 0.5$ (pink circles) and $\alpha = 0.05$ and $\beta = 0.95$ (purple triangles), the individuals with $p \geq 20$ have a very good interaction with their environment; they can always find the point of greatest concentration ($s = 1.0$) and in a very short time ($\tau < 200$), but they cannot coordinate their movement ($v_a \not\approx 1$). Finally, when $\alpha = 0.95$ and $\beta = 0.05$ (green squares), the agents are able to coordinate for $p \geq 40$, and they

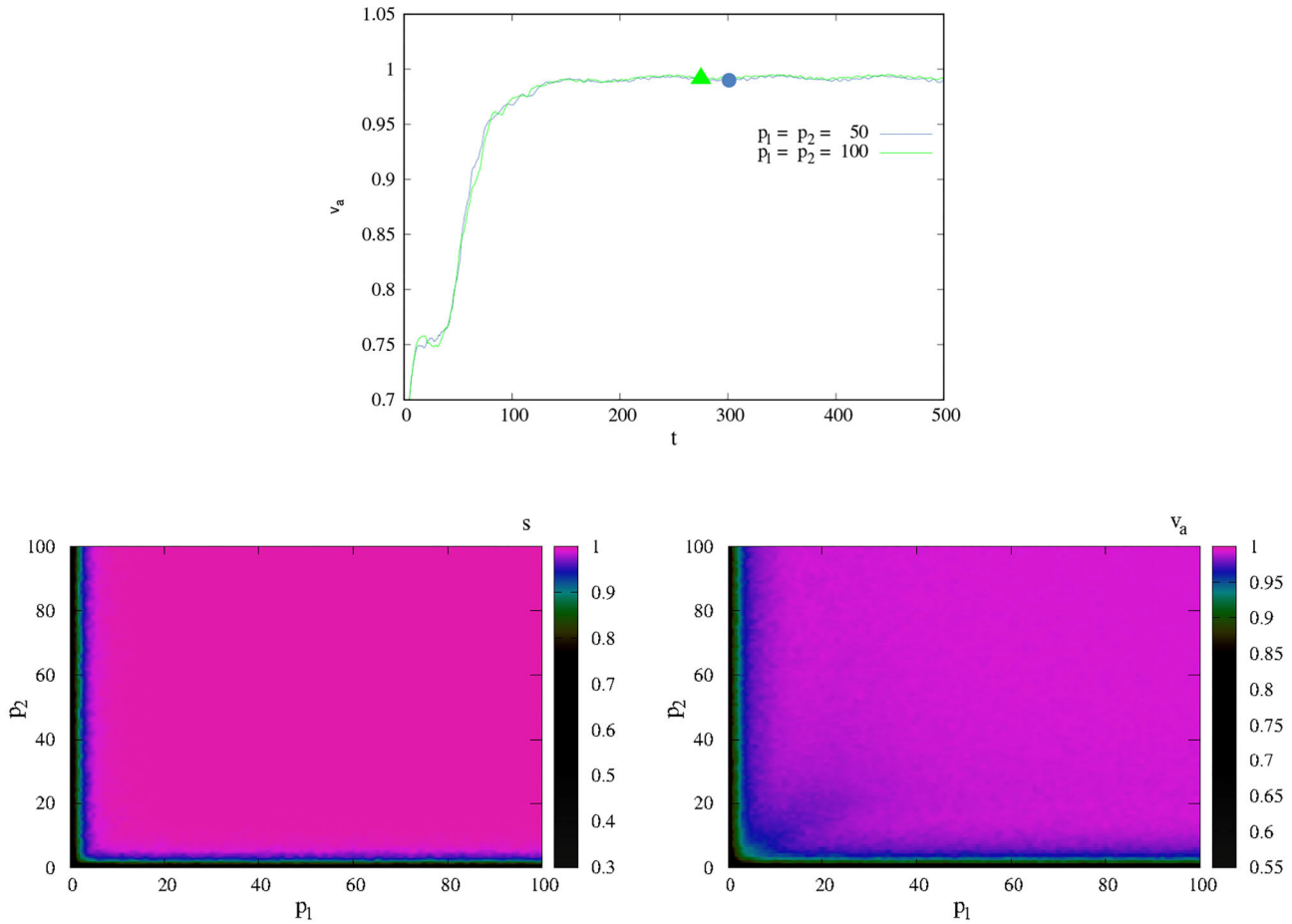

FIGURE 7 | Achieving success in a task. **(Top)** v_a as function of t for different PK values p. Symbols represents the average first-arrival time for $p = 50$ (blue circle) and $p = 100$ (green triangle). **(Bottom)** s (Left) and v_a (Right) as function of (p_1, p_2). In all cases the graphs are the result of averaging 500 different experiments, each one with $t = 2,500$.

can also quickly find the point of greatest concentration ($s = 1.0$ and $\tau < 400$).

The above graphs show us that for medium values of PK p and $\alpha = 0.95, \beta = 0.05$ our SPK agent can have a successful interaction with their environment and coordinating their movement. But to check if they can solve the task correctly, it is necessary to investigate if they arrive at the concentration center in a coordinated way.

Figure 7, top shows the change of v_a as a time increase for SPK agent with different values of PK ($p = 50$-blue line and $p = 100$-green line). We can see that for times >150, the agents are capable of coordination. The blue circle shows the average first-arrival time for agents with p equal to 50, and the green triangle portrays the same quantity but for $p = 100$. Both symbols lie in the section of the curve in which the agents are already coordinated. We can therefore say that for medium, or greater, values of PK ($p \geq 50$), the agents are capable of solving the task successfully.

Figure 7, bottom shows s (Right) and v_a (Left) as functions of (p_1, p_2). For NSA ($p_i \leq 5$), the success rate improves only with the interaction with an SA (pink zone). But for the task to be solved in coordination ($v_a \approx 1$), it is necessary that one of the agents has a medium value of PK ($p_i \geq 40$) and the other has the same or greater p. This means that an agent with high SPK improves the performance of an agent with lower SKP. The PK experience of both agents then arises from their own activity when interacting with others or through their self-other proprioception.

DISCUSSION

In this paper, we addressed the puzzle of proprioception in action from an ESMT and a phenomenological perspective. Arguing that PK coupling cannot be explained solely in terms of a body position sense or in mechanical terms about the pre-programming of the motor outcome, we proposed a theoretical and formal framework to understand how the PK perceptual experience is a form of mastering and dynamical learning about body orientation, possibilities for action, and felt qualitative dynamics. This allows us to take into consideration two missing dimensions in current accounts of proprioceptive perception in action: self-ecological and self-other relationships and felt experiences. Recognizing this type of relational nature has

epistemological implications that can encourage deep research in these issues.

While ESMT has been mostly developed for the visual and tactile modalities, we believe that the arguments and evidence in favor of ESMT should generalize to other perceptive modalities (Lyon, 2014). Here, we have focused on applying this theory to the PK modality. We have presented a minimal model to describe PK-SMCs, which assumes that the perceptual skill or ability of proprioception/kinesthesia is described by a single parameter p. The main model equations portray the fact that proprioception is coupled with kinesthesia, i.e., a proprioceptive agent senses her body and performs it.

Our results showed that NSPK (low p) are not capable of making a distinction between the three sources of variation in the PK sensory signals:

- *PK-SMCs-self:* They cannot recognize their own position in the world, and their movement in it is erratic. This is an immediate consequence of the structure of equations that define the agent's position and movement.
- *PK-SMCs-self-ecological:* Because the NSPK agent are not able to recognize their own position in the world and, therefore, are not capable of moving in it correctly, their interaction with the environment is poor, and they are not capable of recognizing the different signals that come from it. It is impossible for them to solve the task of finding the center of a concentration gradient efficiently.
- *PK-SMCs-self-other:* The impossibility of NSPK agent to recognize their position in the world leads to an impossibility of interacting with another agent. The NSPK is not capable of sensing whether the other is (or is not) inside of their interaction radius.

On other hand, SPK agent (high p) are perfectly capable of making distinctions between the three different sources of variation in PK sensory signals mentioned above. Furthermore, they are capable of solving tasks in coordination with the other, the environment, and both the other and the environment. The PK experience of this kind of agents is constituted by the three PK-SMCs: those that are related to their own orientation and action possibilities in present time or self -proprioception (PK-self); those that arise from their own activity when interacting with the environment or self-ecological-proprioception (PK-self-environment); and those that arise from their own activity when interacting with others or self-other-proprioception (PK-self-other).

A remarkable result is that the agents with medium values of p can make a better distinction between PK-SMs-self, self-ecological, and self-other if they interact with agents with higher values of PK. Interaction helps to improve the performance of the agents. Then, the unit of analysis of ecological and dyadic interaction,—as a minimal form of ecological and social cognition—is thus no longer reduced to the individual, but makes reference to a system as a (self-)organized whole, including the agents involved in the interaction, the process of interaction itself, as well as the ecological context in which these interactions take place.

Despite the minimal PK model's simplicity (or rather thanks to it), this finding might be a good starting point for formalizing Merleau-Ponty's statement that when perceiving others "there exists an internal relation that causes the other to appear as the completion of the system" (Merleau-Ponty, 1945, p. 410). This is because the maintenance of the coordinated behavior, which can take place in two distinct regions of state space depending on whether the agents are jointly moving leftward or rightward, depends on the active participation of the other agent. The proposed distinctions are part of the theoretical and formal approach, but, in reality, these three sources of variation are always intertwined due to felt experiences, perception, and learning, which are ongoing and dynamical processes that in many senses are impossible to consider as separate.

Furthermore, our model shows that this significant increase in the preference for the other agent (with whom it is easiest to coordinate) cannot be explained satisfactorily in terms of only the individual's cognitive assessment of the other's presence: it also requires us to take into account the level of relations between the interactants, as reflected by their capacity for joint contingency recognition and the synchronized timing of their respective assessments. We demonstrate this to be the case in our PK minimal model and thus challenge methodological individualism, as have Kelso et al. (2013)'s coupled dynamical systems and Auvray et al. (2009)'s interactionist account perspectives.

This minimal agent-based model therefore serves as a formal proof of concept that the learning or mastering of skills related to the PK-SMCs-self, PK-SMCs-self-environment, and PK-SMCs self-other, such as when two agents reciprocally participate in the interactive realization of each other's socially contingent actions, is possible in principle. Perhaps in the near future, these findings can also be empirically confirmed in actual psychological experiments of social interaction—in particular those that also take into account the sensorimotor conscious experience of the participants.

In sum, this model is simple and summarizes in a few parameters several mechanisms and actions that could be specified in more explicit ways in a more realistic version. On the other hand, we interpreted the parameters α and β as the capability of an acting agent to make a distinction between the sensory source of her own movement and the sensory source that comes from her interaction with the environment. These are free parameters, and they were adjusted so that the agents could solve a particular task. A possible extension of this minimal PK model would be to consider these parameters α and β as functions of p, which would imply that the capability of an agent to perceive these two kinds of movement sources depended on her ability of PK. Finally, in order to portray the fact that the PK experience is an ability that can be learned (and improved) through experience, a future extension could be that the parameter p changed as a function of time and different kinds of interactions (social and ecological).

A small but growing number of experimental, psychological, and simulation studies have investigated the constitutive role of the ecological and social interaction for proprioception or for

social cognition. Ecological studies about the dynamic touch have begun to produce interesting data. For instance, Asao et al. (2012) demonstrated experimentally that proprioception is important for perceiving the length only through identifying physical invariants and potential movements. In addition, research based on the perceptual crossing paradigm has also contributed to this kind of development. With this aim, Auvray and Rohde (2012) predicts that the acquisition of the ability to detect the responsive presence of others is an embodied skill that goes together with a measurable change in the agent's experience.

However, the potential link between evidence of PK coupling, ESMT, and social interaction is still in need of further development to strengthen its epistemological implications, both because the ESMT of proprioception requires clarification and because its neurophysiological and neuroscientific predictions must be made still more explicit.

CONCLUSIONS

This research prompts us to think not only in reflective terms when we refer to a skilled perceptual PK experience but also on the attentive learning of PK-SMCs and particular kinds of feelings or sensibilities. Nevertheless, from the weak embodiment perspective, it is complicated to extend the neural representation toward peripheral, autonomic, ecological, and social aspects of embodiment. The perspective that we have defended here is a stronger notion of embodiment. We suggest that it is the PK system, with its coupling history of interacting and by the individual's personal experiences, that enables specific perception-action loops, learning to interact and to respond to the world rather than representing it. Specifically, skilled proprioceptive and kinesthetic coupling plays an important role in the felt perceptual experience of spatio-temporal self-orientation in present action and interaction in ways that are irreducible to B-formatted representations.

In our proposed minimal model, the PK perceptual experience of the agents is constituted by three PK-SMCs that are related to its own orientation and action possibilities in present time or self-PK (PK-self); those that arise from its own activity when interacting with the environment or self-ecological-PK (PK-self-environment); and those that arise from its own activity when interacting with others or self-other-proprioception (PK-self-other). Besides helping us to differentiate between NSPK and SPK agent, the model provides important results, including the fact that interaction helps to improve the performance of the agents. This finding might be a good starting point for formalizing the statements of interactions discussed by Merleau-Ponty (1945), Kelso et al. (2013), and Auvray et al. (2009). This minimal agent-based model therefore serves as a formal proof of concept that the learning or mastering of skills related to the different PK-SMCs is possible.

In this sense, PK perceptual experience crystallizes as a specific type of coordination of the organism's action with opportunities afforded by the self, the self-environment, and the self-other-environment. In other words, it is necessary to consider the specific organism-environment interactions that the living process would engage in, tracing the path to overcome the transition from subpersonal representations to personal experience. These abilities are meaningful because the agent has learned them through a history of perception and action coupling and does not require internal comparison models.

We think that the strong embodiment strategy used in this paper, contributes to closing the gap between the content of the proprioceptive-kinesthetic perceptual experience and the skilled possibilities for action.

AUTHOR CONTRIBUTIONS

XG-G provided the original idea. AF-C and GR-F developed the agent-based model. All authors discussed the general outline, the theoretical framework of the article, and contributed to comments and revisions.

FUNDING

This work was supported by Grants PAPIIT-UNAM IT300220, PAPIIT- UNAM IA200720, and Postdoctoral grant DGAPA-UNAM (AF-C), Departamento de Educación. Universidad Iberoamericana, Ciudad de México.

ACKNOWLEDGMENTS

We thank Elmarie Venter and Mateusz Wozniak for their constructive review of this manuscript. AF-C thanks A. Aldana for fruitful discussion and PostDoctoral Scholarship-UNAM for financial support.

REFERENCES

Alaerts, K., Levin, O., and Swinnen, S. P. (2007). Whether feeling or seeing is more accurate depends on tracking direction within the perception-action cycle. *Behav. Brain Res.* 178, 229–234. doi: 10.1016/j.bbr.2006.12.024

Alsmith, A. J. T., and De Vignemont, F. (2012). Embodying the mind and representing the body. *Rev. Philos. Psychol.* 3, 1–13. doi: 10.1007/s13164-012-0085-4

Asao, T., Suzuki, S., and Kotani, K. (2012). "Mechanism of length perception by dynamic touch-proposal of identification-perception model considering proprioception," in *2012 ICME International Conference on Complex Medical Engineering (CME)* (Kobe), 402–407. doi: 10.1109/ICCME.2012.6275734

Asatryan, D., and Feldman, A. (1965). Biophysics of complex systems and mathematical models. functional tuning of nervous system with control of movement or maintenance of a steady posture. I. Mechanographic analysis of the work of the joint on execution of a postural task. *Biophysics* 10, 925–935.

Auvray, M., Lenay, C., and Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas Psychol.* 27, 32–47. doi: 10.1016/j.newideapsych.2007.12.002

Auvray, M., and Rohde, M. (2012). Perceptual crossing: the simplest online paradigm. *Front. Hum. Neurosci.* 6:181. doi: 10.3389/fnhum.2012.00181

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adapt. Behav.* 11, 209–243. doi: 10.1177/10597123031 14001

Beets, I. A., Macé, M., Meesen, R. L., Cuypers, K., Levin, O., and Swinnen, S. P. (2012). Active versus passive training of a complex bimanual task: is prescriptive proprioceptive information sufficient for inducing motor learning? *PLoS ONE* 7:e37687. doi: 10.1371/journal.pone.0037687

Bermúdez, J. L. (2000). *The Paradox of Self-Consciousness*. Cambridge: MIT Press.

Botvinick, M., and Cohen, J. (1998). Rubber hands-feel-touch that eyes see. *Nature* 391, 756–756. doi: 10.1038/35784

Buhrmann, T., Di Paolo, E. A., and Barandiaran, X. (2013). A dynamical systems account of sensorimotor contingencies. *Front. Psychol*. 4:285. doi: 10.3389/fpsyg.2013.00285

Capaday, C., Darling, W. G., Stanek, K., and Van Vreeswijk, C. (2013). Pointing to oneself: active versus passive proprioception revisited and implications for internal models of motor system function. *Exp. Brain Res*. 229, 171–180. doi: 10.1007/s00221-013-3603-4

Cardinali, L., Brozzoli, C., and Farne, A. (2009). Peripersonal space and body schema: two labels for the same concept? *Brain Topogr*. 21, 252–260. doi: 10.1007/s10548-009-0092-7

Connell, L. A., Lincoln, N., and Radford, K. (2008). Somatosensory impairment after stroke: frequency of different deficits and their recovery. *Clin. Rehabil*. 22, 758–767. doi: 10.1177/0269215508090674

Coquery, J.-M., Coulmance, M., and Leron, M.-C. (1972). Modifications des potentiels évoqués corticaux somesthésiques durant des mouvements actifs et passifs chez l'homme. *Electroencephalogr. Clin. Neurophysiol*. 33, 269–276. doi: 10.1016/0013-4694(72)90153-8

Craggs, M., Rothwell, J., and Rushton, D. (1979). Gating of somatosensory evoked potentials by active and passive movements in man [proceedings]. *J. Physiol*. 295:96P.

Crane, T. (1992). *The Contents of Experience: Essays on Perception*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511554582

Crapse, T. B., and Sommer, M. A. (2008). Corollary discharge circuits in the primate brain. *Curr. Opin. Neurobiol*. 18, 552–557. doi: 10.1016/j.conb.2008.09.017

Darling, W. G., Wall, B. M., Coffman, C. R., and Capaday, C. (2018). Pointing to one's moving hand: Putative internal models do not contribute to proprioceptive acuity. *Front. Hum. Neurosci*. 12:177. doi: 10.3389/fnhum.2018.00177

Dayan, E., Casile, A., Levit-Binnun, N., Giese, M. A., Hendler, T., and Flash, T. (2007). Neural representations of kinematic laws of motion: evidence for action-perception coupling. *Proc. Natl. Acad. Sci. U.S.A*. 104, 20582–20587. doi: 10.1073/pnas.0710033104

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci*. 14, 441–447. doi: 10.1016/j.tics.2010.06.009

De Vignemont, F. (2018). *Mind the Body: An Exploration of Bodily Self-Awareness*. Oxford: Oxford University Press. doi: 10.1093/oso/9780198735885.001.0001

Doeden, M. (2014). *Basketball Legends in the Making*. Minnesota: Capstone.

Dreyfus, H. L. (1996). The current relevance of merleau-ponty's phenomenology of embodiment. *Electron. J. Analyt. Philos*. 4, 1–16.

Eklund, G. (1972). General features of vibration-induced effects on balance. *Upsala J. Med. Sci*. 77, 112–124. doi: 10.1517/03009734000000016

Farrer, C., Franck, N., Paillard, J., and Jeannerod, M. (2003). The role of proprioception in action recognition. *Conscious. Cogn*. 12, 609–619. doi: 10.1016/S1053-8100(03)00047-3

Feldman, A. G. (2011). Space and time in the context of equilibrium-point theory. *Wiley Interdiscip. Rev*. 2, 287–304. doi: 10.1002/wcs.108

Feldman, A. G. (2016). "The relationship between postural and movement stability," in *Progress in Motor Control*, eds J. Laczko and M. Latash (Cham: Springer), 105–120. doi: 10.1007/978-3-319-47313-0_6

Fourneret, P., and Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* 36, 1133–1140. doi: 10.1016/S0028-3932(98)00006-2

Fridland, E. (2011). The case for proprioception. *Phenomenol. Cogn. Sci*. 10:521. doi: 10.1007/s11097-011-9217-z

Froese, T., and González-Grandón, X. (2019). How passive is passive listening? Toward a sensorimotor theory of auditory perception. *Phenomenol. Cogn. Sci*. 1–33.

Froese, T., Zapata-Fonseca, L., Leenen, I., and Fossion, R. (2020). The feeling is mutual: clarity of haptics-mediated social perception is not associated with the recognition of the other, only with recognition of each other. *Front. Hum. Neurosci*. 14:560567. doi: 10.3389/fnhum.2020.560567

Gallagher, S. (2003). Bodily self-awareness and object perception. *Theor. Hist. Sci*. 7, 53–68. doi: 10.12775/ths.2003.004

Gallagher, S. (2006). *How the Body Shapes the Mind*. Oxford: Clarendon Press. doi: 10.1093/0199271941.001.0001

Gallagher, S. (2009). Deep and dynamic interaction: response to hanne de jaegher. *Conscious. Cogn*. 18, 547–548. doi: 10.1016/j.concog.2008.12.010

Gallagher, S. (2017). Self-defense: deflecting deflationary and eliminativist critiques of the sense of ownership. *Front. Psychol*. 8:1612. doi: 10.3389/fpsyg.2017.01612

Gallagher, S., and Zahavi, D. (2012). *The Phenomenological Mind*. Abingdon, VA: Routledge. doi: 10.4324/9780203126752

Gapenne, O. (2010). "Kinesthesia and the construction of perceptual objects," in *Enaction: Toward a New Paradigm for Cognitive Science*, eds J. Stewart, O. Gapenne, E. A. Di Paolo (Cambridge: MIT Press), 183–218. doi: 10.7551/mitpress/9780262014601.003.0008

Gapenne, O. (2014). The co-constitution of the self and the world: action and proprioceptive coupling. *Front. Psychol*. 5:594. doi: 10.3389/fpsyg.2014.00594

Gibson, J. (1977). The concept of affordances. *Perceiv. Acting Know*. 1, 67–82.

Goldman, A., and de Vignemont, F. (2009). Is social cognition embodied? *Trends Cogn. Sci*. 13, 154–159. doi: 10.1016/j.tics.2009.01.007

Goldman, A. I. (2012). A moderate approach to embodied cognitive science. *Rev. Philos. Psychol*. 3, 71–88. doi: 10.1007/s13164-012-0089-0

Gonzalez-Grandón, X., and Froese, T. (2018). Grounding 4E cognition in Mexico: introduction to special issue on spotlight on 4E cognition research in Mexico. *SAGE* 26, 189–198. doi: 10.1177/1059712318791633

Goodwin, G., McCloskey, D., and Matthews, P. (1972). The contribution of muscle afferents to kinesthesia shown by vibration induced illusions of movement and by the effects of paralysing joint afferents. *Brain* 95, 705–748. doi: 10.1093/brain/95.4.705

Gordon, J. C., Holt, N. C., Biewener, A. A., and Daley, M. A. (2019). Tuning of feedforward control enables stable muscle force length dynamics after loss of autogenic proprioceptive feedback. *eLife* 9:e53908. doi: 10.7554/eLife.53908

Henri, P. (1902). *La Science et l'hypothése*. Paris: Flammarion.

Hewett, T. E., Paterno, M. V., and Myer, G. D. (2002). Strategies for enhancing proprioception and neuromuscular control of the knee. *Clin. Orthop. Relat. Res*. 402, 76–94. doi: 10.1097/00003086-200209000-00008

Howe, K. A. (2018). Proprioceptive awareness and practical unity. *Teorema* 37, 65–82.

Husserl, E. (1989). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: Second Book Studies in the Phenomenology of Constitution, Vol. 3*. Dordrecht: Springer Science & Business Media. doi: 10.1007/978-94-009-2233-4

Ilmane, N., Sangani, S., and Feldman, A. G. (2013). Corticospinal control strategies underlying voluntary and involuntary wrist movements. *Behav. Brain Res*. 236, 350–358. doi: 10.1016/j.bbr.2012.09.008

Iscla, I., and Blount, P. (2012). Sensing and responding to membrane tension: the bacterial MSCL channel as a model system. *Biophys. J*. 103, 169–174. doi: 10.1016/j.bpj.2012.06.021

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol*. 9, 718–727. doi: 10.1016/S0959-4388(99)00028-8

Kelso, J., Dumas, G., and Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Netw*. 37, 120–131. doi: 10.1016/j.neunet.2012.09.003

Kiefer, A. W., Riley, M. A., Shockley, K., Sitton, C. A., Hewett, T. E., Cummins-Sebree, S., et al. (2013). Lower-limb proprioceptive awareness in professional ballet dancers. *J. Dance Med. Sci*. 17, 126–132. doi: 10.12678/1089-313X.17.3.126

Kimmel, M. (2013). The arc from the body to culture: how affect, proprioception, kinesthesia, and perceptual imagery shape cultural knowledge (and vice versa). *Integr. Rev*. 9:300–348.

Lackner, J. R. (1988). Some proprioceptive influences on the perceptual representation of body shape and orientation. *Brain* 111, 281–297. doi: 10.1093/brain/111.2.281

Lebois, F., Sauvage, P., Py, C., Cardoso, O., Ladoux, B., Hersen, P., and Di Meglio, J.-M. (2012). Locomotion control of caenorhabditis elegans through confinement. *Biophys. J.* 102, 2791–2798. doi: 10.1016/j.bpj.2012.04.051

Lenay, C. (2006). Enaction, externalisme et suppléance perceptive. *Intellectica* 43, 27–52. doi: 10.3406/intel.2006.1326

Lyon, C. (2014). "Beyond vision: extending the scope of a sensorimotor account of perception," in *Contemporary Sensorimotor Theory*, eds J. Bishop and A. Martin (Cham: Springer), 127–136. doi: 10.1007/978-3-319-05107-9_9

McGann, M., and De Jaegher, H. (2009). Self-other contingencies: enacting social perception. *Phenomenol. Cogn. Sci.* 8, 417–437. doi: 10.1007/s11097-009-9141-7

Merleau-Ponty, M. (1945). *Phénoménologie de la Perception*. Paris: Gallimard.

Mitsuo, K., Tomoe, K., Hiroshi, I., Eri, N., Satoru, M., and Toshinori, Y. (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. *Prog. Brain Res.* 142, 171–188. doi: 10.1016/S0079-6123(03)42013-X

Myin, E. (2016). Perception as something we do. *J. Conscious. Stud.* 23, 80–104.

Myin, E. and O'Regan, J. K. (2002). Perceptual consciousness, access to modality and skill theories. a way to naturalize phenomenology? *J. Conscious. Stud.* 9, 27–46.

Nakajima, T., Wasaka, T., Kida, T., Nishimura, Y., Fumoto, M., Sakamoto, M., et al. (2006). Changes in somatosensory evoked potentials and Hoffmann reflexes during fast isometric contraction of foot plantarflexor in humans. *Percept. Motor Skills* 103, 847–860. doi: 10.2466/pms.103.3.847-860

Neisser, U. (1988). Five kinds of self-knowledge. *Philos. Psychol.* 1, 35–59. doi: 10.1080/09515088808572924

Noë, A. (2002). Is the visual world a grand illusion? *J. Conscious. Stud.* 9, 1–12.

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

Noë, A., and O'Regan, J. K. (2000). Perception, attention, and the grand illusion. *Psyche* 6, 6–15.

O'Regan, J. (2011). *Why Red Doesn't Sound Like A Bell: Understanding the Feel of Consciousness*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199775224.001.0001

O'Regan, J. K., Myin, E., and Noë, A. (2004). "Towards an analytic phenomenology: the concepts of "bodiliness" and "grabbiness"," in *Seeing, Thinking and Knowing*, ed A. Carsetti (Dordrecht: Springer), 103–114. doi: 10.1007/1-4020-2081-3_5

O'Regan, J. K., Myin, E., and Noë, A. (2005). Skill, corporality and alerting capacity in an account of sensory consciousness. *Prog. Brain Res.* 150, 55–592. doi: 10.1016/S0079-6123(05)50005-0

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973. doi: 10.1017/S0140525X01000115

O'Regan, J. K., and Noe, A. (2002). "The origin of "feel"," in *From Animals to Animats* (Cambridge) Vol. 7, 27-35.

O'Shaughnessy, B. (1995). "Proprioception and the body image," in *The Body and the Self*, eds J. L. Bermúdez, A. Mercel, and N. Eilan (Cambridge: MIT Press) 175–203.

O'shaughnessy, B. (2008). *The Will: Vol. 1, Dual Aspect Theory*. Cambridge: Cambridge University Press.

Peacocke, C. (2014). *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199699568.001.0001

Philipona, D., O'Regan, J., and Nadal, J. (2003). Is there something out there? Inferring space from sensorimotor dependencies. *Neural Comput.* 15, 2029–2049. doi: 10.1162/089976603322297278

Piaget, J. (1937). *La Construction du réel Chez l'enfant*. París: Delachaux & Niestle.

Prochazka, A. and Ellaway, P. (2012). Sensory systems in the control of movement. *Comprehens. Physiol.* 2, 2615–2627. doi: 10.1002/cphy.c100086

Prosk,e, U. and Gandevia, S. C. (2012). The proprioceptive senses: their roles in signaling body shape, body position and movement, and muscle force. *Physiol. Rev.* 92, 1651–1697. doi: 10.1152/physrev.00048.2011

Sheets-Johnstone, M. (2019). Kinesthesia: an extended critical overview and a beginning phenomenology of learning. *Contin. Philos. Rev.* 52, 143–169. doi: 10.1007/s11007-018-09460-7

Sheets-Johnstone, M. (2020). The lived body. *Human. Psychol.* 48:28. doi: 10.1037/hum0000150

Sherrington, C. S. (1907). On the proprio-ceptive system, especially in its reflex aspect. *Brain* 29, 467–482. doi: 10.1093/brain/29.4.467

Sherrington, C. S. (1918). Observations on the sensual role of the proprioceptive nerve-supply of the extrinsic ocular muscles. *Brain* 41, 332–343. doi: 10.1093/brain/41.3-4.332

Silverman, D. (2018). Bodily skill and internal representation in sensorimotor perception. *Phenomenol. Cogn. Sci.* 17, 157–173. doi: 10.1007/s11097-017-9503-5

Suprak, D. N., Osternig, L. R., van Donkelaar, P., and Karduna, A. R. (2007). Shoulder joint position sense improves with external load. *J. Motor Behav.* 39, 517–525. doi: 10.3200/JMBR.39.6.517-525

Sydney, S. (1996). "Self-knowledge and inner sense," in: *Philosophy and Phenomenological Research LV I*, 249–314.

Thelen, E. (1990). "Coupling perception and action in the development of skill: a dynamic approach," in *Sensory-Motor Organizations and Development in Infancy and Early Childhood*, eds H. Bloch, and B. I. Bertenthal (Dordrecht: Springer), 39–56. doi: 10.1007/978-94-009-2071-2_3

Tsakiris, M. (2015). "The relations between agency and body ownership," in *The Sense of Agency*, P. Haggard and B. Eitam (Oxford: Oxford University Press), 235–256. doi: 10.1093/acprof:oso/9780190267278.003.0010

Varela, F. J. (1999). Present-time consciousness. *J. Conscious. Stud.* 6, 111–140.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press. doi: 10.7551/mitpress/6730.001.0001

Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., and Shochet, O. (1995). Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* 75:1226. doi: 10.1103/PhysRevLett.75.1226

von Holst E and Mittelstaedt, H. (1950). Das reafferenzprinzip. *Naturwissenschaften* 37, 464–476. doi: 10.1007/BF00622503

Warren, W. (2006). The dynamics of perception and action. *Psychol. Rev.* 113, 358–389. doi: 10.1037/0033-295X.113.2.358

Wolpert, D. M., Diedrichsen, J., and Randall, F. (2011). Principles of sensorimotor learning. *Nat. Rev. Neurosci.* 12, 739–751. doi: 10.1038/nrn3112

Wolpert, D. M., and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nat. Neurosci.* 3, 1212–1217. doi: 10.1038/81497

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882. doi: 10.1126/science.7569931

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 25 January 2021
doi: 10.3389/fpsyg.2020.611827

From Affective Arrangements to Affective Milieus

*Paul Schuetze**

Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

In this paper, I develop the concept of *affective milieus* by building on the recently established notion of *affective arrangements*. Affective arrangements bring together the more analytical research of situated affectivity with affect studies informed by cultural theory. As such, this concept takes a step past the usual synchronic understanding of situatedness toward an understanding of the social, dynamic, historical, and cultural situatedness of individuals in relation to situated affectivity. However, I argue that affective arrangements remain too narrow in their scope of analysis since their focus mainly lies on local, marked-off, and unique constellations of affect relations. They neglect the more mundane and day-to-day affect dynamics of social life. Hence, I introduce the notion of affective milieus, which brings to light the everyday, ubiquitous affective engagements of individuals with their socio-material surroundings. Affective milieus specifically call attention to how commonplace affect relations create territories in the social universe which form and mold individuals all the time. In that way, this paper apprehends and advances recent developments in the research on situated affectivity.

Keywords: situated affectivity, affect, affective arrangements, milieus, situatedness, social space, cultural affect

OPEN ACCESS

Edited by:
Leon De Bruin,
Radboud University Nijmegen, Netherlands

Reviewed by:
Laura Candiotto,
Free University of Berlin, Germany
Gerhard Thonhauser,
Darmstadt University of Technology, Germany

***Correspondence:**
Paul Schuetze
paul.schuetze@outlook.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *29 September 2020*
Accepted: *31 December 2020*
Published: *25 January 2021*

Citation:
Schuetze P (2021) From Affective Arrangements to Affective Milieus. Front. Psychol. 11:611827. doi: 10.3389/fpsyg.2020.611827

INTRODUCTION

In recent years, the research on *situated affectivity* has led to the insight that affective phenomena should not only be analyzed in isolation, as marked-off individual mental states or happenings. Instead, researchers agree that these phenomena need to be addressed as being situated, as manifested in the interactions of agents and their surroundings (see e.g., Griffiths and Scarantino, 2009; Gallagher, 2013; Krueger, 2014; Colombetti and Krueger, 2015; Stephan and Walter, 2020). Advancing this endeavor, in the field of situated affectivity, the concept of *affective arrangements* was proposed as a theoretical tool to reveal that the intimate effects affectivity has on the interactional dynamics within socio-material settings (Slaby et al., 2019).[1] This concept takes a step past the usual synchronic understanding of situatedness and goes beyond a focus on singular emotions, moods, existential feelings, or sentiments. It focuses on dynamic situatedness, on "local constellations of elements that give rise to specific relational domains of affecting and being affected" (Slaby et al., 2019, p. 5). Affective arrangements capture ensembles of persons, things, discourses, spaces, and behaviors in which *affect* is a unique modulator – they describe material-discursive formations orchestrated in compositions of particular *affect*

[1]Since the concept presented in this paper, i.e. affective milieus, builds on affective arrangements, it is significantly different from the one of affective scaffolding discussed by Colombetti and Krueger, 2015. This is because they have a different analytical focus, for details on this difference see e.g. Slaby, 2016.

relations (Slaby et al., 2019, p. 5). As such, they combine the socio-material situatedness of individuals and their affective relationality; and they emphasize that affect dynamics largely unfold between multiple actors in social domains of practice (Slaby, 2019b, p. 60). This makes possible an understanding of the social and cultural situatedness of individuals in relation to affectivity; and thereby, affective arrangements bring together the more analytical studies of affectivity and emotion (e.g., Griffiths and Scarantino, 2009; Stephan et al., 2014; Colombetti and Krueger, 2015), and affect studies informed by cultural theory (e.g., Gregg and Seigworth, 2010).

Within the current debate, this, of course, takes situated views of affectivity for granted (see e.g., Gallagher, 2013; Colombetti, 2018; Stephan and Walter, 2020); and basic assumptions from this field are presupposed, most importantly that "there is no pre-formed, independently existing individual that comes into a pre-formed, independently existing world[...]. Rather, it is the environment and the individual which together determine who and what they are" (Stephan and Walter, 2020, p. 15). Building on these assumptions, in the following, I survey the concept of affective arrangements in more detail, illustrating its core characteristics. However, I argue that affective arrangements remain too narrow. As I will elaborate in the Affective Arrangements section, by focusing on local and specific situations, they only capture special kinds of marked-off arrangements and address only very particular affect relations. Even though affective arrangements enable an understanding of the social situatedness of individuals in terms of affect relations in the first place, they neglect a societal and more large-scale view on situated affectivity. Thus, I take the theoretical concept of affective arrangements as an outset, and I apply its central ideas to the societal level. In doing so, I introduce the notion of *affective milieus*.[2] Crucially, in contrast to affective arrangements, the concept of affective milieus is not attached to unique and local ensembles, but it brings to light acculturated and situated modes of being in general.[3] It calls attention to a person's habitualized affective engagements with her socio-material surroundings, how this relationality shapes her entire mode of being, not only in idiosyncratic and demarcated situations, and how this engagement manifests in particular spaces of the social world.

The first half of this paper will be concerned with laying out the conceptual framework of affective arrangements. I start with an introduction to the idea of *relational affect*, a central aspect of affective arrangements. Secondly, I introduce the concept of affective arrangements and analyze it in more detail. Then, I move beyond the concept of affective arrangements by pointing out its shortcomings while still upholding its main ideas. The second half of this paper develops the notion of affective milieus. Building upon the essential features of affective arrangements, I apply the key insights to a more societal and large-scale view. Finally, I illustrate the significance of affective milieus with a concrete example making clear the advantages this concept brings with it.

AFFECTIVE ARRANGEMENTS

Relational Affect

The notion of affective arrangements builds on an understanding of affect as a relational phenomenon, as not being restricted to individual agents, but as being a dynamic between bodies of various kinds (Slaby, 2019b, p. 61). The following provides a short introduction to this relational conception of affect, making way for a more detailed analysis of affective arrangements. The idea of relational affect takes a perspective on affectivity which focuses on situatedness and relationality, i.e., on the material and ideational relations unfolding "between […] 'bodies' whose potentialities and tendencies are thereby continuously modulated in mutual interplay" (Slaby et al., 2019, p. 4). To get a grip on this idea, take the following example: Suppose you are sitting at a restaurant table with some friends. You are loosely talking, arguing, and laughing while eating, drinking, and simply being there with each other. As is often the case in these situations, you may feel inclined to lean toward one side of the table and engage with this side more than the other, or you may only want to talk with the person sitting right across from you. Sometimes, it is even the case that the seating order already determines how the evening will go, how you will experience the atmosphere, how long you will want to stay, and how enjoyable the conversations will be. With any person leaving or joining the table, the whole situation can change; what was an intimate conversation may turn into shallow small talk, or what was a boring back and forth may suddenly become exciting. Even more subtle factors, like you having a drink in front of you or not, might affect how the evening evolves.

Intuitively, one will recognize all of these more or less subtle experiences and sensations. These are prime examples of the affect relations unfolding between social and material bodies. Yet, it will be difficult to put a finger on them, to specify what they truly are, because they are not graspable in terms of "clearly demarcated mental states" (Slaby, 2019b, p. 61). Rather, they are subtle changes in the relational dynamics between a person and her surroundings which influence how she experiences a situation and engages in it. These are the particularities which the notion of relational affect brings into focus.

In that sense, affect "is construed as a relational dynamic between individuals and in situations – a dynamic that is prior to individual experience, even, in a sense, prior to the individual subject as such" (Slaby, 2019b, p. 60). While "affectivity" denotes the general capacity of a person to be sensitive to, and affected by, what matters to her, "affect" characterizes the

[2]Merleau-Ponty also talks about "affective milieus" (e.g., Merleau-Ponty, 1945/2012, p. 156). But, different from the meaning pursued in the current paper, for Merleau-Ponty, an affective milieu manifests around an individual's body. For him, an affective milieu is "the sector of our experience that clearly has sense and reality only for us" (Merleau-Ponty, 1945/2012, p. 156). As such, an affective milieu denotes the surroundings which affectively matter to the individual; these affective relations bring into existence these surroundings for the individual body in the first place (Merleau-Ponty, 1945/2012, p. 140; see also Roald et al., 2018). Since I am building on the notion of affective arrangements, I do not take up Mearleau-Ponty in the following.

[3]By "mode of being" I touch upon Heideggerian terminology, i.e., being-in-the-world. By different "modes of being" I, therefore, refer to different modes of being-in-the-world (see Wheeler, 2018; Thonhauser, 2020).

concrete relations in which a most basic form of affectivity substantiates. This relational idea of affect goes back to Baruch Spinoza's complex metaphysical framework of substance monism (for a detailed discussion of this background, see Mühlhoff, 2018; Slaby and Mühlhoff, 2019). Without subscribing to the whole conceptual landscape of Spinoza, the essential point here is to recognize affect as a relational phenomenon which is constitutive of the individual subject (Seyfert, 2012; Mühlhoff, 2018). Affect, thus understood, might be viewed in terms of *relational affect dynamics* which express how a subject is situated in the world, i.e., in its social and physical surroundings (Mühlhoff, 2018, p. 20). This entails a "radically relational and dynamical understanding of individuals" which are grasped as "transiently stabilizing node[s] in an encompassing relational dynamic" (Slaby and Mühlhoff, 2019, p. 30). Individuals are constantly entangled in ways of affecting and being affected and they always have to be understood in terms of this relational fabric (Mühlhoff, 2018, p. 50). More specifically, "[t]he individual gets constituted processually … in a network of affective relatedness" (Mühlhoff, 2015, p. 1013). In that way, a focus on relational affect brings with it a developmental constructivist analysis of subjects (Slaby et al., 2019, p. 5). And so, the notion of affect puts emphasis on the base layer or the substructure on which the experiences, feelings, ultimately the individual subject itself is built upon (Åhäll and Gregory, 2015, p. 5). With this idea of relational affect in mind, the next section introduces and analyses the concept of *affective arrangements*.

Affective Arrangements and Their Conceptual Background

Affective arrangements are first and foremost a theoretical tool to shine light onto local sociomaterial settings in which unique affect dynamics emerge and are continuously modified (Slaby et al., 2019, p. 3). In the following, I first provide an overview of the theoretical structure and the background of affective arrangements and then move on to a clarifying example.

Affective arrangements owe their name and their principal theoretical origin to the concept of *agencements* developed by Deleuze and Guattari (Deleuze and Guattari, 1987; Slaby et al., 2019).[4] In their work, Deleuze and Guattari describe agencements as heterogeneous ensembles which consist of different artificial and natural components (Deleuze, 2006, p. 179). An agencement is a co-functioning unity which is defined in terms of the relations between its integrated elements; together these elements are laid out in an orchestrated, specific, and coherent whole in which they work together for a certain amount of time (Müller, 2015, p. 28). Yet, an agencement does not have an essence in and of itself, but it is entirely reliant on the relations of its elements, on the way, these elements are connected and work together coherently (Nail, 2017, p. 23). In an agencement, vastly different elements come together, and despite their difference, they portray a form of consistency, they create a unique identity and claim a territory in which the agencement persists (Wise, 2005, p. 77). In short, "an *agencement* is a fragmentary, open-textured formation: a concatenation of components that keep their distinctness" while still working together as a whole (Slaby et al., 2019, p. 6). To underpin this abstract idea with an intuitive picture, one may think of an agencement as a "dry-stone wall" (Deleuze and Guattari, 1994, p. 23). The individual elements, the stones, are not added and glued into a homogenous whole, rather they retain their individuality while still being part of a unity, a heterogeneous arrangement which works together as a whole, as a dry-stone wall.

However, since agencements only form the theoretical basis for the concept of affective arrangements, there are still differences. As the name already implies, affective arrangements put particular focus on affect relations, or more specifically, relational affect is their very basis. Going back to the dry-stone wall, one may say that affective arrangements are exactly that, fragmentary formations which form an orchestrated whole in virtue of their relatedness (Slaby, 2019a, p. 110). And, crucially, the relations between the elements are affect relations, i.e., affect is the glue which holds the stone wall together and prevents the stones from falling left and right. In that way, affect relations are the core of affective arrangements, they connect all elements within an arrangement, such that the arrangement becomes a unity demarcated from its surroundings (Slaby et al., 2019, p. 6). Following the concept of Deleuze's and Guattari's agencements, we can thus say that affective arrangements are ensembles of "persons, things, artifacts, spaces, discourses, behaviors, expressions, or other materials that coalesce into a coordinated formation of mutual affecting and being-affected" (Slaby, 2019a, p. 109).

Another defining precursor concept to affective arrangements is Foucault's *dispositif* (Foucault, 1980; Slaby, 2019a, p. 109). Just like an agencement, the dispositif denotes a heterogenous ensemble consisting of various elements, such as discourses, institutions, laws, scientific statements, and philosophical and moral propositions, which are connected *via* their relations (Foucault, 1980, p. 194). And similarly, a dispositif describes the network of relations between the various elements. But other than agencements, a dispositif specifically highlights the social and political power structures that come with it (Seyfert, 2012, pp. 33–34). In that way, the notion of a dispositif captures the "strategies of relations of forces supporting, and supported by, types of knowledge," (Foucault, 1980, p. 196) and as such it frames the setting in which certain things can be said, whereas others cannot – in which certain things can be conceived, whereas others cannot (Foucault, 1980, p. 194). What the idea of a dispositif adds to the texture of affective arrangements are the strategic power relations that manifest in ensembles of affect dynamics. Within an affective arrangement, the integrated elements take on specific roles, which are only partly due to their individuality, but which are largely the result of the relational framing of the respective formation (Slaby et al., 2019, p. 8). Moreover, a dispositif is defined by "a certain kind of genesis" (Foucault, 1980, p. 195). This means the network of relations of forces has a historicity – it always describes a particular way of becoming, of how it emerged and stabilized. Affective arrangements portray the same historicity. They are never just there, but "they emerge out of multiple formative trajectories,

[4]The most common translation, retained by Brian Massumi, of *agencement* would be *assemblage*. But, as this brings with it various semantic problems, I make use of the original term (see e.g., Phillips, 2006; Buchanan, 2015).

for example, histories of fine-tuning, of combining and recombining of components" (Slaby et al., 2019, p. 8). There is a genesis to affective arrangements manifested in the histories of the affect relations between its components, in habituated ways of affecting and being affected, and in the acculturation of rules, discourses, spaces, expressions, and other materials.

Summarizing the above, we may adhere that affective arrangements are heterogenous ensembles of natural and artificial elements, in which local patterns of affect dynamics form a unique affective texture. Such idiosyncratic formations are held together by specific affect relations; they prompt new affect dynamics, but also modify and guide them. Integrated individuals are subject to mechanistic relations, as they take on affective roles and acculturate modes of being, and by processes of habituation they become part of a functioning whole orchestrated by affect relations (Slaby, 2019a, p. 116).

To provide a clarifying example, consider a family gathering at Christmas. Parents, grandparents, children, aunts and uncles, cousins, and other relatives meet for their annual Christmas dinner. There are the classic tree, the candles, the Christmas smells, the typical food, and some presents on the side. All of this takes place at the same location each year, in the grandparents' house. Importantly, this recurring event creates the same overall affective atmosphere: the feeling of Christmas. It is a historically grown tradition, which the family members have acculturated, and each new member, such as a new partner or a newborn child, is readily integrated. This illustrates the performative open-endedness of affective arrangements (Slaby, 2019a, p. 110). They are not pre-determined and rigid constellations, rather they possess a dynamic openness, in the sense that affective arrangements are "capable of expanding into their surroundings by incorporating new elements" (Slaby, 2019a, p. 110). Within the Christmas dinner there are natural and artificial elements integrated, and they come together to form a unity, a functioning whole. Each component, be it family member, tree, or present, retains its individuality, but takes on a role and becomes part of a network of relations creating the Christmas dinner. Much the same as a dry-stone wall, the Christmas dinner is a heterogenous ensemble consisting of distinct components which nevertheless cohere and create a unity held together by affect relations.

As the dinner carries on, some of the family will still be eating while the kids might already be finished and have left the table to play. By then, others will be in the kitchen, washing the dishes and preparing dessert. There are various interactions taking place simultaneously at different locations: the usual talk at the table, the more private conversation in the kitchen and the untamed play of the kids. While the overall pattern of the arrangement persists, it is constantly changing and transforming (Slaby, 2019a, p. 111). The different family members all have slightly different experiences, depending on their point of view and the people they are engaged with. Each member takes on a different role, which they have habituated over the years before, and which is strategically placed within the overall ensemble. One overall affective atmosphere has different yet similar segments, depending on the different affective interactions and relations. And all of this depends not only on the synchronic happenings, but also on the multi-track historicity of the affective arrangement, namely on the particular family history, traditions, and relationships, but also on "gender roles, cultural habits and commonsense behavioral expectations" (Slaby et al., 2019, p. 7). It is exactly this, the unity despite the situational diversity, the uniqueness of exactly this arrangement emerging from particular histories which come together, and the dynamic stabilization by processes of relational co-constitution, which is captured by an affective arrangement (Slaby, 2018a, pp. 209–210; Slaby et al., 2019, p. 7).

Going Beyond Affective Arrangements

Having clarified the theoretical background, I now focus on aspects of affective arrangements which provide a starting point for introducing the larger-scale concept of affective milieus. An important point concerns the way individuals are seamlessly integrated in affective arrangements, how they attach to and are influenced by local affect-generating and co-constituting set-ups (Slaby, 2018a, p. 210; Slaby et al., 2019, p. 7). Most of the time, individuals automatically fit into the arrangement, they appropriately engage in the various interactions and become part of the whole by conforming to the overall pattern. Consider the Christmas dinner: Every family member behaves and feels according to the Christmas-like structure, according to the particular interaction partners (e.g., children or grandparents) and according to their specific location (e.g., kitchen or dinner table). In that way, even though there usually is no strongly felt pressure to abide to particular norms, each individual is integrated and acts according to a role (e.g., von Maur, 2018, p. 100). This means that mostly without noticing, without force, and mainly without actively being restricted in their individuality, all individuals being part of an affective arrangement behave and experience according to a role. In this way, the perspective of affective arrangements reveals the manner in which "subtle forms of a reciprocal affective interplay" produce and enforce entire modes of being (Slaby et al., 2019, p. 7). With reference to Foucault's dispositif, this highlights the subliminal, yet influential, power relations which are at play in these networks of affecting and being affected.

Nonetheless, there are contrary situations in which the structures within an affective arrangement do not remain opaque, and the modulating plays of power are strongly felt. In the dinner example, suppose that one person at the table might start to make questionable jokes and comments. Commonly, there is an implicit rule to ignore these comments and not to make a big deal out of it for the sake of peace, so to speak. However, other guests might not want to let these comments be expressed unnoticed. For them, the affective arrangement is in tension with personal commitments, their roles within the overall formation deeply conflict with their individual identity. In those situations, norms of interaction are strongly felt, they appear at the surface and individuals are no longer seamlessly integrated. They notice how the situation binds them to a particular behavior, yet they want to act against it. Breaking one of these tacit norms in such situations requires effort, for it interrupts the fluidity of the situation and often causes irritation. The background nature

of the affective arrangement will get lost and the affective atmosphere will possibly change. Here, it is important to note that such instances are less common compared to the situations in which one does not notice the underlying structures. Often times, individuals are unaware of the affective arrangements they are in, and so, for the most part affective arrangements seamlessly integrate individuals. But, most importantly, such tense situations emphasize the underlying force of affective arrangements. The effort it costs to deviate from the implicit rules and from the appropriate behavior or feeling (e.g., not laughing at the inappropriate jokes) indicates the force with which an affective arrangement usually incorporates individuals.[5] From such set-ups of modulating and constituting affective relationality emerge new modes of being – "the individual subject … is … a complex 'product' of the sustained modulation by affect-intensive social domains" (Slaby, 2016, p. 2). Going back to Foucault's dispositif, this makes explicit how the relations within an affective arrangement are relations of power. There are norms, ideas, and rules concretely embodied in these relations (Slaby, 2016, p. 8), such that individuals within the arrangement are subject to these power dynamics merely by being integrated in the arrangement – often being unaware of it (Slaby et al., 2019, p. 5).

Another essential aspect of affective arrangements is that they do not appear just like that, but they are the result of congregating histories, they are historically grown. The Christmas dinner is not realized from one day to another, but it has some kind of a genesis. This includes cultural trajectories, such as gender roles, behavioral norms, and other material-discursive processes, as well as the family and individual history. When these various lines meet and intersect, their concurrence creates the Christmas dinner. Such a multi-track historicity makes affective arrangements "'conservation devices' in which histories of interaction and of collective habituation have become sedimented" (Slaby et al., 2019, p. 7). This means that affect dynamics within affective arrangements necessarily rely on processes of becoming and on devices of acculturation, bringing to life a sedimented past (Slaby et al., 2019, p. 9). This particular emphasis takes a step away from a synchronic conception of situatedness. Instead it moves toward a complex, temporal, and diachronic comprehension of situated affectivity. In this way, affective arrangements acknowledge and provide a grip on the subtle, powerful, and intimate influences of affectivity – on the affective and "ontogenic dynamics that are formative of subjects" (Slaby et al., 2019, p. 7).

However, as mentioned in the beginning, I argue that affective arrangements still remain too narrow. They are geared toward picking out local settings of affect relations with a focus on idiosyncrasy. Not every family dinner is an arrangement, but only the ones that stick, the ones with a historicity, with a distinctiveness which makes people resonate (Slaby et al., 2019, pp. 5, 8–9). This means that the focus lies on singular and exceptional formations, such as the Christmas dinner. In order for this to happen, particular trajectories have to intersect and affect dynamics have to stabilize, forming "a unique local patterning of relational affect, giving shape to a potentially idiosyncratic affective texture or formation inherent in a specific place at a time" (Slaby, 2019a, p. 116). But despite this emphasis on local restriction and idiosyncrasy, affective arrangements tell us that situated affect dynamics are diachronic processes of becoming, that they orient and modify subjects, and that they form their entire mode of being. Evidently, this is not restricted to unique situations or formations such as the Christmas dinner, rather it is an everyday mechanism, repeatedly recurring. This is where affective arrangements are too limited in their scope. Such ontogenic processes do not just remain within physically restricted or "cranky" circumstances (Slaby et al., 2019, p. 8), but they subsist, they are there all the time. This is why the next section introduces the concept of *affective milieus*, taking seriously a more large-scale and wholistic view on situated affectivity. As such, affective milieus focus on *everydayness*: They bring to light the power relations manifested in the affect dynamics of social life, and they reveal the affective formative processes subjects are exposed to and immersed in every day.

AFFECTIVE MILIEUS

Affect Dynamics as Orientation Devices

The theoretical tools of affective arrangements make apparent the concrete ways in which subjects are constituted by day-to-day affect dynamics. In order to further illustrate these fundamental mechanisms I bring to mind insights from the field of *critical phenomenology*. These approaches help to bridge the gap between the more localized analysis of affect dynamics, i.e., affective arrangements, and a large-scale, societal level viewpoint, i.e., affective milieus. Without building on the whole conceptual landscape of critical phenomenology, I focus on the rationale that historical and social structures "play a constitutive role in shaping the meaning and manner of our experience" (Guenther, 2020, p. 12). Essentially, this brings with it a large-scale analysis of the "social structures that make our experience of the world possible and meaningful" (Guenther, 2020, p. 15). Although these approaches are not explicitly developed in connection to affectivity, they nicely translate to the affective realm; and even tough they usually focus on the contingency of our experiences, they also shine light on general processes of subjectification. Take, for instance, the work of Sara Ahmed in "*Queer Phenomenology*" (Ahmed, 2006). In her approach, Ahmed describes the implications of what it means to be *oriented* in the world. While this idea is not concretely developed in terms of affect, it, nonetheless, helps to show that affect relations are powerful orientation devices in everyday interactions and not only in marked-off arrangements. Therefore, this work makes the transition from affective arrangements to affective

[5]One may think of situations in which individuals feel strongly restricted by the norms at play. Take for instance the gender norms present in a very traditional Christmas dinner (e.g., women preparing the food and men talking at the table). Despite the imperative objections one should have when encountering such norms, the essential point remains: an individual's passive integration into the arrangement. Norms are not explicitly set up – no one openly states these gender norms before every Christmas dinner – rather they subliminally guide individuals.

milieus genuinely explicit. Now, while some of these insights are already implicitly present in the concept of affective arrangements, Ahmed's detailed visualizations help to make them concrete.

Then, what does it mean to be oriented in the world? A person's orientation determines what is close to her, and what is distant. Metaphorically speaking, things that are close are in sight, and things that are distant are out of sight. Different orientations limit what a person can do, what she can experience and what she may think. Ahmed starts with the simple example of sitting at a desk. Sitting and facing the desk implies a certain orientation in this very moment. Things on the table are near and in reach, things in the background are out of sight and out of reach. Being oriented toward the desk brings into focus specific matters. For instance, the work on the desk is of primary importance while the background, such as the family sitting in the kitchen, is of less interest. In that sense, the orientation toward the desk shapes what a person experiences as close or distant, as important or negligible, as doable or unfeasible (Ahmed, 2006, pp. 25–65).

Although, the concept of orientation is rather abstract, and Ahmed devotes large parts of her book to it, for the current purpose, it suffices to connect this idea to the study of situated affectivity. The very physical access Ahmed provides can be abstracted and applied in a figurative manner. To give a simple example from the realm of affectivity consider affective atmospheres (e.g., Riedel, 2019). A bright and sunny winter day with blue skies affects people in an entirely different way than a stormy, gray, rainy, and cold winter day. Both days have a very distinct atmosphere which people are embedded and entrenched in. Depending on this atmosphere, people may be oriented in a specific way, some things may be in sight, while others may remain hidden. For instance, on the sunny and bright day people may feel inclined to go outside, do exercise, or clean their apartment. Whereas on the rainy and gray day, they may want to be lazy, stay inside, watch a movie, or read a book. As Ahmed says: "What is reachable is determined precisely by orientations that we have already taken. Some objects do not even become objects of perception…: they are 'beyond the horizon' of the body, and thus out of reach" (Ahmed, 2006, p. 55). This example illustrates how Ahmed's generic concept of orientation can be understood in connection to affective phenomena.

Subsequent to these ordinary examples, Ahmed points out that the concept of orientations also applies on a more fundamental and sustained level. It is not only in some situations that orientations are influential, but a person's very mode of being is constituted by habituated orientations. Take again the Christmas dinner mentioned above: During the dinner, there are different roles for each family member, and these roles come with a specific orientation and ability to navigate. The socio-material affect dynamics composing the dinner orient and align all individuals in a certain way. As a consequence, the family implements implicit orientations without noticing and without force. This is enforced by the whole family: Unwittingly, every action and every word facilitate these orientations. With a nod to Deleuze's and Guattari's agencements, we can say that in the Christmas dinner, ensembles of trajectories come together aligning people in a peculiar way. Yet, importantly, such ensembles do not just develop in the face of the moment, but they are acculturated over years. And so, the dinner comes about as an arrangement of historically interwoven lines subjecting and habituating individuals to a unique material-discursive structure, enforcing orientations which outlast the moment and stick with the individuals over time (Ahmed, 2006, pp. 79–92).

However, as already mentioned above, such processes of habituation are not only present in unique situations, such as the Christmas dinner. Rather, we are always subjected to lines shaping and directing our orientations – "persistent social structures influence our capacity to experience the world, not just in isolated instances but in a way that is deeply constitutive of who we are and how we make sense of things" (Guenther, 2020, p. 13). In terms of relational affect, this reveals that affect dynamics act upon individuals all the time, they do not only orient individuals in certain atmospheres, such as a sunny or rainy day, or in specific situations, such as the Christmas dinner.

In the context of racism, specifically of *racializing perception*, Al-Saji (2014) provides a concrete example of how these processes unfold. Without going into too much detail here, Al-Saji analyzes how sedimentation and habituation manifested in affect relations tacitly constitute our visual processes such that certain things "*cannot be seen otherwise*" (Al-Saji, 2014, p. 138). As she states, "I can see bodies as raced only because I cannot see them otherwise," (Al-Saji, 2014, p. 139) and this is deeply rooted in the habituated structures of our vision, as well as in acculturated affect relations configuring this vision. Now, while Al-Saji focuses on racialized bodies and racializing perception, at its core, her account is similar to what Ahmed captures with orientations. And so, this translates into a more general claim: "What is 'otherwise' is not only occluded from vision, but also from feeling, imagination, and understanding" (Al-Saji, 2014, p. 141). By analyzing the structures of vision, Al-Saji shows us how "habituated and socialized affects" form individuals in general – even primal processes, such as perception, are fundamentally constituted by patterns of affect relations (Al-Saji, 2014, p. 140). In other words, just as "habits of seeing owe to a social, cultural, and historical field," (Al-Saji, 2014, p. 138) entire modes of being are the product of historically sedimented patterns of affecting and being affected (see also von Maur, 2018, pp. 224–232).

In short, the above connection to critical phenomenology highlights that affect dynamics form individuals, not only within affective arrangements but, more importantly, also in day-to-day dealings. All the affect relations a person is exposed to come together as transformative patterns of affecting and being affected, whereby they permanently constitute the person and how she makes sense of things. This means that affect dynamics are fundamental mechanisms of acculturation enforcing particular modes of being: They diachronically form what individuals find within reach, what they can see, what they can feel, and what they can think or imagine. This diachronic formation of subjects cannot be captured within the concept of affective arrangements, which remains within the limits of analyzing

localized and idiosyncratic ensembles. Here, the notion of *affective milieus* makes a start.

From Affective Arrangements to Affective Milieus

Generally, affective milieus inherit the core features of affective arrangements. Affective milieus are forms of agencements: They are heterogenous formations of natural and artificial elements which are held together by affect dynamics, but which do not have an essence in themselves; rather they are defined only in terms of these relational dynamics (Nail, 2017). They have a multi-track historicity as various trajectories intersect within them, and so they function as conservation devices preserving, molding and generating affect relations. Affective arrangements are like agencements "not simply a happenstance collocation of people, materials, and actions" (Buchanan, 2015, p. 385), but "a *specific* tangle of relations of affecting and being affected" (Slaby et al., 2019, p. 6). In the same way, affective milieus are specific tangles of everyday affect relations. Affective milieus also share the relational forces and plays of power captured by a dispositif. They are material-discursive ensembles developing, blocking, enforcing, and stabilizing power relations which support and are supported by types of knowledge (Foucault, 1980, p. 196). As such, they manifest in a network of forces held up by particular affect dynamics, and thus they subject integrated individuals to distinct power relations. In that way, affective milieus share the core features of affective arrangements, such as the ones adopted from agencements and dispositifs.

The crucial difference to affective arrangements is that affective milieus do not describe locally marked-off situations or ensembles which stabilize once in a while, which need to resonate with individuals, or which lure them into their positions (Slaby et al., 2019, p. 9). In a sense, affective milieus are always there: They do not usually have an attracting character, but they are structures residing in social domains of practice. They are *societal and large-scale* formations, which subdivide social space in a way that individuals are seamlessly integrated simply by being there. In contrast to affective arrangements, affective milieus are not cranky or strange compositions, they are not something extraordinary, and they are not something purposeful (Buchanan, 2015, p. 385; Slaby et al., 2019, p. 8). Affective milieus describe the day-to-day affect dynamics individuals are immersed in; they capture commonplace affect relations and identify them as powerful orientation devices, "as … process[es] of domestication – of making some objects and not others available" (Ahmed, 2006, p. 117).

This means that detached from affective arrangement, the spatial openness and the everydayness of affective milieus put focus on the permanent subjectification effects of affect dynamics. Affect relations are at the heart of a "material-discursive subject constitution … [which] … is a matter of effective framing and re-molding of subjectivity and selfhood" (Slaby, 2016, p. 7). It is exactly this aspect which affective milieus take up, as they shine light on the impact of large-scale societal formations. Adopting the notion of affective milieus highlights that "the subject is an active, environmentally embedded, and affectively situated agent" (Piredda and Candiotto, 2019, p. 136). As we have seen above in the digression into critical phenomenology, subjects are not only shaped by processes in unique localized situations, such as affective arrangements. But, the entire subject is constantly changing and building itself through ways of affecting and being affected (Piredda and Candiotto, 2019, p. 139). In other words, "every past experience of being-in-relation … shapes and forms the present and future individual potential" of the subject (Mühlhoff, 2015, p. 1013).[6] This fundamental embeddedness in social space is picked out by affective milieus, and it is the dimension which marks the major difference to affective arrangements. In other words, the concept of affective milieus allows us to take a step back and get a grip on the various locally unbound affect relations which form an individual. This perspective goes beyond a selective focus on work environments, public transports, sports games, shopping malls, and other local settings (cf. Slaby et al., 2019, p. 9). Rather, this new angle of view puts emphasis on the multifaceted and ubiquitous affect relations coming together in a subject.

To clarify the difference between affective milieus and affective arrangements, take the following example: Suppose a person going home after a demonstration. The concept of an affective arrangement captures the particular dynamics of the demonstration; but once the demonstration is over, once this specific formation dissolves and the person detaches from the arrangement, the notion of an affective arrangement loses its grip. While the person leaves behind the particular affective arrangement, this does not necessarily entail leaving behind all of the affect dynamics or the orientations that were present within the arrangement. Instead, particular significance relationships might still remain with her, and particular dynamics may transfer to other areas of her life as well. For instance, when meeting friends after having participated in a rally, the topics of discussion will likely evolve around similar subjects; or when making certain decisions, the just experienced orientations will still remain influential. In that way, the affect dynamics live on in the individual. And so, these dynamics function as ongoing orientation devices bringing some things into sight, while making others impossible to see.

Of course, this does not happen immediately, merely by going to a single demonstration. But individuals are subject to infinitely many affect dynamics, not all of which are parts of affective arrangements such as demonstrations, but which might just be parts of daily routines, interactions, or other processes. These affect dynamics all come together in the individual; they do not suddenly vanish, nor can the individual simply detach from them. They move the individual, they stick with them, they embed them in ensembles of affect dynamics and relations, and thus they make up their lifeworld. Such dynamics do not remain singular points of contact, but they

[6]Here, it is important to note that subjects can actively change their interaction with the environment, they can partly change the ways in which they affect and are being affected. This way, affective practices have a vital transformative character (see Candiotto, 2019; Piredda and Candiotto, 2019). However, in the current paper, I cannot take up this implication as I employ a descriptive approach particularly focusing on the substantial influences of already existing affect relations (see also Slaby, 2016).

are part of a whole – they are parts of formations in the material and social life of individuals. They come together as a network of affect relations, meshed together, always transforming, stabilizing, modulating, and producing ways of affecting and being affecting. Ultimately, these affect dynamics constitute the individual. Such locally spread everyday dynamics are neglected by affective arrangements, and they are revealed by affective milieus.

Affective Milieus

As shown above, although affective milieus inherit the core features of affective arrangements, there are some key differences. We have already seen that affective milieus are forms of agencements and forms of dispositifs. However, in contrast to agencements as taken up by affective arrangements, affective milieus are not highly localized, idiosyncratic structures with a mechanistic function. Rather, they are social formations which are there all the time. They reside in day-to-day socio-material relations and in the daily affective interactions of individuals. In that regard, affective milieus share the characteristics of a dispositif, as they are manifested in the relations between various elements on a social scale. Yet, different from a dispositif, affective milieus are composed of affect relations which function as the glue holding the various elements together. Moreover, milieus do not always have a "major function at a given historical moment" (Foucault, 1980, p. 195), as Foucault points out regarding dispositifs. Affective milieus can be without a historical function or a specific purpose. In essence, they are formations in social space, which individuals are always already situated within.

I mentioned before that affective milieus are large-scale societal formations. This means that in contrast to affective arrangements, they describe enduring ensembles of natural and artificial elements which are not restricted to a local setting. Yet, just as an agencement, affective milieus create a territory (Wise, 2005, p. 78). The affect dynamics composing an affective milieu occupy a certain space, they demarcate an area in which the milieu persists. Individuals integrated in these particular affect dynamics inhabit this territory, they are embedded in it such that the individual and the socio-material environment mutually constitute each other (Slaby, 2018b, pp. 331–332). Importantly, the territories so created have to be understood as abstract formations within the social space, as spaces defined by particular ways of affecting and being affected within social domains. This means that affective milieus do not literally delimit a marked-off physical territory. Rather, they demarcate a space in the social world understood as "a multidimensional system of co-ordinates" (Bourdieu, 1985, p. 724). In that way, affective milieus share core features with the spatial idea of a *social group*. Just as members of a social group "have a specific affinity with one another because of their similar experience or way of life" (Young, 2004, p. 43), elements of an affective milieu are connected to each other by similar affect relations and modes of being. As Iris Marion Young describes, social groups "are not entities that exist apart from individuals, but neither are they merely arbitrary classifications of individuals" (Young, 2004, p. 44). Similarly, affective milieus do not exist apart from their elements and the network of relations between them. It follows that an affective milieu only exists in virtue of shared and interconnected social and material relations of bodies; a specific milieu is not always there, but it has a history of stabilizing dynamics. This also means that these dynamics constantly change. The elements and the affective milieu constitute each other – as the elements change, the milieu changes and as the milieu changes, the elements change. Just like social groups are "fluid" constellations as "they come into being and may fade away" (Young, 2004, p. 47), affective milieus are constantly changing and transforming ensembles. Of course, these are not rapid transformations, but they entail a longer lasting development – a process of domestication and habituation, making some socio-material bodies and not others available by changing the affect relations between bodies over time.

At this point, it is important to note that the current paper merely gets a grip on the formations of affective milieus. In a next step it needs to be analyzed how these structures can be transformed, how they are not merely conservation devices, but possibly also vehicles of change. In this regard, further research may provide promising contributions, explicating how affective milieus can be altered and how individuals can change patterns of affecting and being affected. In fact, existing research on the transformative impact, especially of affective practices already offers fruitful insights into these questions (e.g., von Maur, 2018, ch. 5; Candiotto, 2019; Piredda and Candiotto, 2019). Once more, this highlights the unique perspective that comes with an analysis of affect dynamics in regards to societal issues: By its very nature it already provides access to avenues of change and to perspectives of rearrangement. And so, the concept of affective milieus not only presents itself as a descriptive tool but also offers space for transformative beginnings.

Now, the affiliation with the concept of a social group together with the perspective of societal change emphasizes the scale of affective milieus. Namely, they demarcate networks of affect relations on a *societal level*. As such, the concept of affective milieus functions similarly to the one of a social group; it arranges the social space into different formations. However, it is important to stress once again that affective milieus are composed of heterogenous elements. They are restricted neither to social nor to material relations, but they combine both. The linkage to social groups merely illustrates the scale on which affective milieus operate. Just as there are different social groups and classes, there are different affective milieus coexisting. This comparison makes concrete that an affective milieu is essentially a societal scale formation, subdividing the social universe into different territories, manifested in locally unbound affect relations.

Naturally, there are no sharp, clear-cut distinctions or absolute breaks in the social world (Bourdieu, 1987, p. 13). Therefore, affective milieus share aspects of what Pierre Bourdieu describes in the context of *social classes*. Social classes are overlapping, bordering upon one another with gradual borders, just as the "boundaries of a cloud or a forest" (Bourdieu, 1987, p. 13). The boundaries of a social class "can thus be conceived of as lines or as imaginary planes, such that the density (of the

trees or of the water vapor) is higher on the one side and lower on the other" (Bourdieu, 1987, p. 13). The boundaries of an affective milieu are exactly the same. Just as a person can be right in the dense center of a forest, she can be in the intense, strongly integrated part of an affective milieu; and just as she can be at the light edge of the forest, where the forest gradually meets the meadow, she can be at the less intense edge of an affective milieu, where one milieu meets and passes into another. Social classes structure the social space and so do affective milieus. Abstractly speaking, a person is assigned an area within "the social universe" in virtue of a multitude of variables that apply to her, and this location attributes her to a certain class (Bourdieu, 1987, p. 4). In a similar manner, a person is located in the territory of a certain affective milieu depending on the affect relations she is embedded in.

Affective milieus demarcate territories in social space. These territories are almost like habitats for the integrated elements, they are socio-material environment these elements live in. Crucially, these territories are demarcated by particular affect dynamics where certain trajectories intersect and where unique ways of affecting and being affected are at play. By their very nature and by being formations within the social universe, affective milieus are not rigid, but fluid structures which are always transforming; and they are not clear-cut unities but dynamically open ensembles which are marked-off from their surroundings by gradual borders. All of the integrated individuals are similarly oriented and share a similar horizon, depending on their place within the affective milieu. This gives rise to an ensemble of elements, almost like a collective involving "shared orientations toward and around objects … [which] … would be an effect of the repetition of this direction over time" (Ahmed, 2006, p. 118).

To clarify the above, let us take an example and concretely apply the idea of an affective milieu. Suppose the cluster of environmentally conscious people, or more broadly speaking the *eco dispositif* if you will. The concept of an affective milieu allows us to frame this formation in terms of situated affectivity and grasp this formation as an *affective eco milieu*. This means that we can delineate a territory in social space where very particular socio-material relations are at play. For instance, the eco milieu may comprise individuals who share the same concerns, such as how to reduce plastic or CO_2 emissions, or who have similar subjects to discuss with family and friends, for instance, how to buy more sustainable products. This territory may also be characterized by particular groups and specific activities, for example, individuals may come together and share their interest in gardening or farming. Moreover, this milieu also includes material relations such as owning sustainable clothes or foods, which are bought for instance from wholefood shops. And it may even be manifested in different kinds of work, as individuals may want to be doing something good for the world by choosing a workplace that accords with their principles. In short, there are very particular socio-material dynamics which compose the eco milieu. At its center, this milieu is a dense formation knitted by unique ways of affecting and being affected, and it gradually fades toward its edges, where the affect relations are loose-knit, where they overlap and intersect with bordering milieus. Depending on how strongly an individual is integrated and involved in the respective dynamics, it is located in the dense center or the lighter edges.

On the one hand, this example indicates that different individuals can be situated in different, even contrasting affective milieus. This would result simply from being involved in different affect dynamics. In the next section, I will go into more detail regarding this issue. On the other hand, this example also shows that there is a sort of unity and connection among the individuals integrated in the same affective milieu, simply because they are arranged in a shared network of relations which brings them together. They are part of a heterogenous formation, in which each person has her own life while still moving around the same socio-material settings as the others. Importantly, this example pinpoints the difference between affective arrangements and affective milieus by highlighting that individuals can meet in the same arrangements while being in a different milieu. Take for instance the Christmas dinner and the eco milieu. In one affective arrangement, the Christmas dinner, there may be individuals who are embedded in vastly different affective milieus, for instance, the eco milieu vs. an opposing milieu. And so, in contrast to affective arrangements, affective milieus can be described by a rather broad range of generic features, such as people sharing concerns; people engaging in similar topics; people exchanging and discussing with others the subjects that affect them, in families or with friends, at work or in their sports group; people reading or hearing the news and reacting in certain ways; and people buying and consuming similar media and other goods. Of course, this is only a small number of the dynamics which make up an affective milieu. Yet, they are examples of the concrete affect relations constituting affective milieus.

OUTLOOK: FUTURE PERSPECTIVES

To show the significance of the concept of affective milieus, I bring to mind the topic of climate change. As an exemplary instance of this topic, I want to focus on the public debate about the sustainability of cars. This example will purposefully be exaggerated and I am well aware that there are more subtle undertones which I deliberately pass over. Yet, with this hyperbolic juxtaposition, I hope to pointedly contour the issues at stake, and to specifically highlight the unique understanding that comes with the notion of affective milieus. On the one side of the exemplary debate about the sustainability of cars, environmental activists demand that owning and driving cars ought to be more expensive to meet the actual costs of emitting an excessive amount of CO_2 through individual transport. This should be achieved, for example, by introducing carbon taxes that would make gas more expensive. The contrary position – the car lobbyist – usually stresses the cultural and practical value of cars in addition to important social unrests that might result from higher gas prices (see e.g., the Yellow Vests movement). These two positions strongly oppose each other

and whenever there is something to be done in either direction, reactions of the opposing side are harsh.

Framing this in terms of affective milieus, it becomes clear why both sides oppose each other so strongly. Usually, the wish for cars to be more expensive comes from younger people, often people (e.g., students and young families) who live in cities, where living without a car is rather easy. Moving within their environment is dominated by public transport, bikes, short distances to the supermarkets, and most places are within reach. Cars are even perceived as a burden for them. The streets are occupied by parking spots stealing valuable space within the city. Cars are loud and dirty, and they are making life among them unpleasant. Cyclists and pedestrians encounter cars as dangerous objects, almost as living entities which anonymously pass by accompanied by an aura of discomfort and fear. There are very particular affective relations such people have and do not have in connection to cars. Hence, they can easily conceive of a life without a car, and they may even enjoy the idea of a car-free city. Additionally, these people are immersed in very peculiar affect dynamics: They engage in certain activities, they might, for example, seek to escape the city by attending a small garden; they usually meet like-minded people who navigate in similar settings, and share the same work or living situation; they only consume particular things, e.g., exclusively buying environmentally friendly clothes and organic food.

The opposing side is often represented by people who own, love, and need a car. As such, they use a car more frequently, for instance to get to work or because they live in more rural areas. Their world is characterized by driving a lot, by long distances and spending a lot of time in or near their cars. They see an aesthetic value in owning a car. And so, a car is not simply a car, but an object of desire. This object should have certain favorable and appealing characteristics. For instance, owing an SUV in a city has no practical value at all, but it brings with it a peculiar feeling. And so, the affect relations these people have with and around their cars are vastly different to the ones described before. And similarly, these people are involved in their very own affect dynamics: Consumption priorities are different, e.g., their car has a high personal value, it signals their social status, motivating them to hold and spend their money accordingly; their social contacts largely evolve around people who also own cars and can only be reached by car, or with whom they go on trips and vacation. In contrast to the other camp, their areas of life are shaped less by environmental concerns, i.e., the activities these people are engaged in are not so much focused on environmental friendliness. They might for instance carelessly do winter sports or fly to vacation destinations, the clothes and food they buy might not be sustainably produced.

Each of the two camps is situated in an affective milieu with its peculiar socio-material dynamics. The people in each milieu are oriented in very different directions (although not always in such a contrasting manner). Very particular things come into reach and become possible when being oriented around a world involving cars or around a world without cars. This also means that within such an affective milieu only a limited set of solutions comes into sight when approaching a problem. For either of the two camps, it requires a lot of effort to see and comprehend the ideas and thoughts of the other side. This is simply because such ideas and thoughts are not within reach from the affective milieu they themselves are situated in. Relating back to the Christmas dinner example, it requires work to not just let the ignorant jokes slide at the table. One needs to step out of given norms and break with one's habituated mode of being. In a similar way, individuals within the affective milieu of "liberal car-related people" need to step out of their habituated being in order to bring other solutions into reach.

Relating this to the broader example of climate change, we can see how the study of situated affectivity can contribute to the analysis of such issues. The concept of an affective milieu makes this contribution concrete. It is not enough to present people with new data in order for them to change their behavior, or their way of life more generally. It is not even enough to present them with the concerns of other people. For a person to look beyond her affective milieu, she needs to be aware of the specific power relation she is embedded in. Relating back to the field of critical phenomenology, the goal then needs to be to create possibilities of reflecting and changing one's relations with the world.

AUTHOR CONTRIBUTIONS

The author confirms sole responsibility for the conception, development and composition of this article.

ACKNOWLEDGMENTS

I would like to thank Jan Slaby, Achim Stephan, Lia Nordmann, and the members of the Reading Club of the Institute of Cognitive Science at University of Osnabrück for extensive, valuable, and critical feedbacks. I also thank the two reviewers for their very detailed and constructive comments.

REFERENCES

Åhäll, L., and Gregory, T. (eds.) (2015). "Introduction: mapping emotions, politics and war" in *Emotions, politics and war* (London and New York: Routledge).

Ahmed, S. (2006). *Queer phenomenology.* Durham, London: Duke University Press.

Al-Saji, A. (2014). "A phenomenology of hesitation" in *Living Alterities. Phenomenology, embodiment and race*. ed. E. S. Lee (Albany, NY: University of New York Press), 133–172.

Bourdieu, P. (1985). The social space and the genesis of groups. *Theory Soc.* 14, 723–744.

Bourdieu, P. (1987). What makes a social class? On the theoretical and practical existence of groups. *Berk. J. Sociol.* 32, 1–17.

Buchanan, I. (2015). Assemblage theory and its discontents. *Deleuze Stud.* 9, 382–392. doi: 10.3366/dls.2015.0193

Candiotto, L. (ed.) (2019). "Emotions in-between: the affective dimension of participatory sense-making" in *The value of emotions for knowledge* (Cham: Palgrave Macmillan).

Colombetti, G. (2018). "Enacting affectivity" in *The Oxford handbook of 4E cognition*. eds. A. Newen, L. D. Bruin and S. Gallagher (Oxford: Oxford University Press), 571–588.

Colombetti, G., and Krueger, J. (2015). Scaffoldings of the affective mind. *Philos. Psychol.* 28, 1157–1176. doi: 10.1080/09515089.2014.976334

Deleuze, G. (2006) *Two regimes of madness*. ed. D. Lapoujade (Cambridge: MIT Press).

Deleuze, G., and Guattari, F. (1987). *A thousand plateaus capitalism and schizophrenia*. (Trans. B. Massumi) Minneapolis and London: University of Minnesota Press.

Deleuze, G., and Guattari, F. (1994). *What is philosophy?* New York: Columbia University Press.

Foucault, M. (1980). "The confession of the flesh" in *Power/knowledge. Selected interviews and other writings, 1972–1977*. ed. C. Gordon (New York: Pantheon Books), 194–228.

Gallagher, S. (2013). The socially extended mind. *Cogn. Syst. Res.* 25, 4–12. doi: 10.1016/j.cogsys.2013.03.008

Gregg, M., and Seigworth, G. J. (eds.) (2010). *The affect theory reader*. Durham, London: Duke University Press.

Griffiths, P., and Scarantino, A. (2009). "Emotions in the wild" in *The Cambridge handbook of situated cognition*. ed. P. Robbins (Cambridge: Cambridge University Press), 437–453.

Guenther, L. (2020). "Critical phenomenology" in *50 concepts for a critical phenomenology*. eds. G. Weiss, A. V. Murphy and G. Salamon (US: Northwestern University Press), 11–16.

Krueger, J. (2014). Varieties of extended emotions. *Phenomenol. Cogn. Sci.* 13, 533–555. doi: 10.1007/s11097-014-9363-1

Merleau-Ponty, M. (1945/2012). *Phenomenology of perception*. (Trans. D. A. Landes) London and New York: Routledge.

Mühlhoff, R. (2015). Affective resonance and social interaction. *Phenomenol. Cogn. Sci.* 14, 1–19. doi: 10.1007/s11097-014-9394-7

Mühlhoff, R. (2018). *Immersive Macht: Affekttheorie nach Spinoza und Foucault*. New York: Campus Verlag.

Müller, M. (2015). Assemblages and actor-networks: rethinking socio-material power, politics and space. *Geogr. Compass* 9, 27–41. doi: 10.1111/gec3.12192

Nail, T. (2017). What is an assemblage? *SubStance* 46, 21–37.

Phillips, J. (2006). Agencement/Assemblage. *Theory Cult. Soc.* 23, 108–109. doi: 10.1177/026327640602300219

Piredda, G., and Candiotto, L. (2019). The affectively extended self: a pragmatist approach. *Humana.Mente* 12, 121–145.

Riedel, F. (2019). "Atmosphere" in *Affective societies. Key concepts*. eds. J. Slaby and C. von Scheve (London and New York: Routledge), 85–95.

Roald, T., Levin, K., and Køppe, S. (2018). Affective incarnations: Maurice Merleau-Ponty's challenge to bodily theories of emotion. *J. Theor. Philos. Psychol.* 38, 205–218. doi: 10.1037/teo0000101

Seyfert, R. (2012). Beyond personal feelings and collective emotions: toward a theory of social affect. *Theory Cult. Soc.* 29, 27–46. doi: 10.1177/0263276412438591

Slaby, J. (2016). Mind invasion: situated affectivity and the corporate life hack. *Front. Psychol.* 7:266. doi: 10.3389/fpsyg.2016.00266

Slaby, J. (2018a). Affective arrangements and Disclosive postures. *Phänomenologische Forschungen* 2, 197–216. doi: 10.28937/1000108209

Slaby, J. (2018b). "Existenzielle Gefühle und In-der-Welt-sein" in *Emotionen. Ein interdisziplinäres Handbuch*. eds. H. Kappelhoff, J. -H. Bakels, C. Schmitt and H. Lehmann (Stuttgart: Metzler), 326–339.

Slaby, J. (2019a). "Affective arrangements" in *Affective societies: Key concepts*. eds. J. Slaby and C. von Scheve (London and New York: Routledge), 109–118.

Slaby, J. (2019b). "Relational affect: perspectives from philosophy and cultural studies" in *How to do things with affects*. eds. E. van Alphen and T. Jirsa (Leiden: Brill), 59–81.

Slaby, J., and Mühlhoff, R. (2019). "Affect" in *Affective societies: Key concepts*. eds. J. Slaby and C. von Scheve (London and New York: Routledge), 27–41.

Slaby, J., Mühlhoff, R., and Wüschner, P. (2019). Affective arrangements. *Emot. Rev.* 11, 2–12. doi: 10.1177/1754073917722214

Stephan, A., and Walter, S. (2020). "Situated affectivity" in *The Routledge handbook of phenomenology of emotions*. eds. T. Szanto and H. Landweer (Abingdon and New York: Routledge).

Stephan, A., Wilutzky, W., and Walter, S. (2014). Emotions beyond brain and body. *Philos. Psychol.* 27, 65–81. doi: 10.1080/09515089.2013.828376

Thonhauser, G. (2020). "Martin Heidegger and Otto Friedrich Bollnow" in *The Routlegde handbook of phenomenology of emotions*. eds. T. Szanto and H. Landweer (Abingdon and New York: Routledge).

von Maur, I. (2018). Die Epistemische Relevanz des Fühlens. Osnabrück. Available at: https://nbn-resolving.org/urn:nbn:de:gbv:700-20180807502 (Accessed December 2020).

Wheeler, M. (2018). "Martin Heidegger" in *The Stanford Encyclopedia of Philosophy*. ed. E. N. Zalta (Stanford University). Available at: https://plato.stanford.edu/archives/fall2020/entries/heidegger/ (Accessed July 2020).

Wise, J. M. (2005). "Assemblage" in *Gilles Deleuze: Key concepts*. ed. C. J. Stivale (Montreal and Kingston: McGill-Queen's University Press), 77–88.

Young, I. M. (2004). "Five faces of oppression" in *Oppression, privilege, and resistance: Theoretical perspectives on racism, sexism, and Heterosexism*. eds. L. M. Heldke and P. O'Conor (Boston: McGraw-Hill), 37–63.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

PERSPECTIVE
published: 29 January 2021
doi: 10.3389/fpsyg.2021.543076

Toward an Embodied, Embedded Predictive Processing Account

*Elmarie Venter**

Institute for Philosophy II, Ruhr University Bochum, Bochum, Germany

In this paper, I argue for an embodied, embedded approach to predictive processing and thus align the framework with situated cognition. The recent popularity of theories conceiving of the brain as a predictive organ has given rise to two broad camps in the literature that I call *free energy enactivism* and *cognitivist predictive processing*. The two approaches vary in scope and methodology. The scope of *cognitivist predictive processing* is narrow and restricts cognition to brain processes and structures; it does not consider the body-beyond-brain and the environment as constituents of cognitive processes. *Free energy enactivism*, on the other hand, includes all self-organizing systems that minimize free energy (including non-living systems) and thus does not offer any unique explanations for more complex cognitive phenomena that are unique to human cognition. Furthermore, because of its strong commitment to the mind-life continuity thesis, it does not provide an explanation of what distinguishes more sophisticated cognitive systems from simple systems. The account that I develop in this paper rejects both of these radical extremes. Instead, I propose a compromise that highlights the necessary components of predictive processing by making use of a mechanistic methodology of explanation. The starting point of the argument in this paper is that despite the interchangeable use of the terms, prediction error minimization and the free energy principle are not identical. But this distinction does not need to disrupt the *status quo* of the literature if we consider an alternative approach: Embodied, Embedded Predictive Processing (EEPP). EEPP accommodates the free energy principle, as argued for by free energy enactivism, but it also allows for mental representations in its explanation of cognition. Furthermore, EEPP explains how prediction error minimization is realized but, unlike cognitivist PP, it allocates a constitutive role to the body in cognition. Despite highlighting concerns regarding cognitivist PP, I do not wish to discredit the role of the neural domain or representations as free energy enactivism does. Neural structures and processes undeniably contribute to the minimization of prediction error but the role of the body is equally important. On my account, prediction error minimization and free energy minimization are deeply dependent on the body of an agent, such that the body-beyond-brain plays a *constitutive* role in cognitive processing. I suggest that the body plays three constitutive roles in prediction error minimization: The body *regulates* cognitive activity, ensuring that cognition and action are intricately linked. The body acts as *distributor* in the sense that it carries some of the cognitive load by fulfilling the function of minimizing prediction error. Finally, the body serves to *constrain* the information that is processed by an agent. In fulfilling these three roles, the agent and environment enter into a bidirectional relation through influencing and modeling the structure of the other. This connects EEPP to the

OPEN ACCESS

Edited by:
Leon De Bruin,
Radboud University Nijmegen, Netherlands

Reviewed by:
Gunnar Declerck,
University of Technology Compiegne, France
Michael David Kirchhoff,
University of Wollongong, Australia

***Correspondence:**
Elmarie Venter
elmarie.venter@rub.de;
elmarie.venter@ruhr-uni-bochum.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *24 March 2020*
Accepted: *12 January 2021*
Published: *29 January 2021*

Citation:
Venter E (2021) Toward an Embodied, Embedded Predictive Processing Account. Front. Psychol. 12:543076. doi: 10.3389/fpsyg.2021.543076

free energy principle because the whole embodied agent minimizes free energy in virtue of being a model of its econiche. This grants the body a constitutive role as part of the collection of mechanisms that minimize prediction error and free energy. The body can only fulfill its role when embedded in an environment, of which it is a model. In this sense, EEPP offers the most promising alternative to cognitivist predictive processing and free energy enactivism.

Keywords: predictive processing, embodiment, mechanistic explanation, free energy, prediction error

INTRODUCTION

This paper defends an embodied approach to the predictive processing framework that is aligned with the broader setting of situated cognition. Inspired by principles in biological sciences and computer sciences, the predictive processing framework (henceforth, PP) has gained much popularity in cognitive science in recent years. This account of cognition turns the traditional account of cognition upside down: instead of the brain gathering information about the world, processing information, and then employing it in the output of action, the brain is constantly making predictions about the world. The account has been applied to explain a variety of processes in the brain, and aims to provide a unifying perspective of perception, action and cognition. This is agreed upon by most researchers in the field but the exact relationship between perception, action and cognition remains a contested topic in the literature on PP (Colombo and Wright, 2016). The surprising number of varied interpretations of PP may lead one to question whether they, in fact, refer to the same idea. The aim of this paper is to investigate this question and offer an embodied approach to PP (Clark, 2016; Kirchhoff, 2017, 2018; Kirchhoff and Kiverstein, 2019). I do this by differentiating between two popular interpretations—cognitivist PP and free energy enactivism—and then carving out an account most compatible with a strong embodied account of cognition. I take strong embodiment to mean that both neural structures and wider bodily structures *constitute* cognitive processes insofar as the body not only contributes to (or enables) the function of the predictive system (to minimize prediction error) but also directly fulfills this function without mediation by mental representations (Shapiro, 2004; Rowlands, 2010). This is contrasted with weaker embodiment claims which take cognitive processes to be *dependent* (to varying degrees) on bodily structures and processes (Rupert, 2009; Alsmith and de Vignemont, 2012).

The paper is organized as follows. I briefly describe the grounding principles of PP and focus on highlighting the distinction between the free energy principle and prediction error minimization[1]. The free energy principle sets out to explain all mind and life, and is typically applied to explaining why dynamic systems avoid disorder or dispersal (Friston, 2013a; Sims, 2016). Given the wide scope of the free energy principle, I set out to narrow down the discussion to the cognitive domain by presenting the relevant features of PP in terms of prediction error minimization. I then investigate two interpretations of PP: cognitivist PP leans toward a commitment to internalism, and free energy enactivism undertakes the task to explain dynamic, coupled engagement with the world. After critically examining the scope and explanatory ambitions of these two interpretations, I defend a mechanistic explanation of PP and use this as a starting point to develop an embodied account of PP. On the mechanistic approach, all components of the system that realize the function of the system are important and must be included in the explanation. Following this, I argue that the body be granted a strong constitutive role in an explanation of cognition because it fulfills the function of prediction error minimization without necessarily being mediated by mental representations.

SETTING THE SCENE

The objective of this section is to provide a bird's eye view of the necessary features of predictive processing (PP). This section is intentionally vague given that more specific features will be discussed in the subsequent sections. What I wish to highlight is the distinction between the free energy principle and prediction error minimization. Though the two concepts are difficult to separate and often used interchangeably in the literature, they make different predictions and vary in scope and application (Bruineberg et al., 2018; Hohwy, 2020). Any description of PP starts with an understanding of the free energy principle which is defined as follows: "any self-organizing system that is at equilibrium with its environment must minimize free energy" (Friston, 2010) where free energy refers to a state associated with disorder or uncertainty. The principle is based on the fact that biological systems have a limited range of states in which they can survive. It is therefore necessary for an organism to maintain itself within its possible range of states by minimizing disorder and uncertainty; failure to do so leads to dispersal and ultimately death. The idea upon which PP is build is that in order for a system to maintain itself within a particular range of states, it requires the capacity to predict future states. In sophisticated systems, like human agents, this means tracking and representing the causes of sensory states. This process is realized by generative models with different sets of priors about

[1]The divorce of the free energy principle from predictive processing is becoming more popular and several recent papers argue for such a separation. Hohwy (2020), for example, argues that the free energy principle offers a normative theory that is a mathematical and conceptual analysis whereas predictive processing is a falsifiable process-theory. Bruineberg et al. (2018) also argue for a conceptual distinction between free energy minimization and prediction error minimization. Although they also propose that the two concepts are incompatible whereas Hohwy and myself do not.

the environment and the agent. The primary function of these generative models is to maintain a set of hypotheses about the world that generates the most accurate predictions of the incoming information and consequently minimize uncertainty about the environment (free energy). Free energy is evaluated using two factors: an agent's sensory states and a recognition density (i.e., the aforementioned probabilistic representation of the hidden causes of sensory states) (Friston, 2010). Free energy minimization is a principle of optimization that can be applied at many different levels of analysis and at different timescales, explaining how we maintain bodily states such as, for example, blood sugar levels (Seth, 2013) to how we maintain an optimal narrative model of ourselves (Hohwy and Michael, 2017), and even explaining social cognition by means of interoceptive inference (Fotopoulou and Tsakiris, 2017). The use of "generative model" is cautiously applied and does not *necessarily* imply contentful representation because it can be applied beyond the neural domain. When the free energy principle is applied to the neural domain, the amount of free energy is calculated as the sum of all prediction error which is defined as the divergence between the probability distribution encoding the sensory states and the recognition density. It is interpreted as the mismatch between what is predicted and the incoming sensory stimuli.

In the neural domain, PP is defined by the idea that processing stimuli is driven by top-down processes. This is commonly referred to as prediction error minimization (Hohwy, 2013). To see a structured world is to use existing generative models of the world to shape a virtual version of sensory perturbations from the top down. Thus rather than reconstructing the world, the system is "constantly trying to *guess the present*" (Clark, 2017b p. 727, emphasis in original). Generative models are constantly updated so that the best possible top-down predictions are generated to meet bottom-up transmissions. Better top-down predictions mean that more incoming information is matched and explained away (which results in less uncertainty). The process of "explaining away" incoming information leaves only prediction error to be propagated within the system. This bidirectional process occurs at different spatial and temporal scales operating at many different levels of a processing hierarchy where, at each level, the system is trying to predict its own sensory states. The important feature in this schema is interaction between the different levels, where higher level predictions involve more abstract and temporally extended states and lower-level predictions process more fine-grained states, such as lines, edges and textures of surfaces. The predictions that pervasively determine perceptual experiences are extracted from higher levels and prior knowledge based on statistical estimation. Statistical estimation refers to a calculation of accuracy and precision within a range of likely and probable predictions that explain sensory causes. The function of the whole system is for top-down predictions to meet the incoming signals and become more successful at making predictions about the world. Estimates at each level in the hierarchy are also predictive of each other in order to assist with the successful execution of this function. Thus, prediction is not just from one level to the next but also occurs between models at a single level. This strategy is efficient in that it minimizes computing power because mismatches between top-down and bottom-up information only update generative models which already exist (Metzinger and Wiese, 2017).

Prediction error minimization is the main objective of the system (the brain is commonly the system referred to in this context). Predictions can be accurate or inaccurate to varying degrees. There is a direct correlation between the accuracy of a prediction and how well fitted a generative model is in that an accurate prediction is an indication of a successful generative model. If the prediction is accurate, nothing more needs to be done and the generative model is accurate with respect to the state of the world. If bottom-up signals are not accurately predicted, the mismatched information is transmitted as prediction error until the model (more or less) matches the state of the world. Prediction error can be minimized in two ways: perceptual inference and active inference. Perceptual inference involves model revision based on prediction errors. Prediction errors are transmitted up the hierarchy and the generative model is updated. Active inference is a process in which the agent acts upon, or changes, the world in order to bring about the state of the world predicted by the current best generative model. It can be argued that active inference can be explained in entirely internalist terms insofar as predictions about bodily movements and its causes on the environment is an inferential process. Cognitivist PP is committed to the view that active inference is a result of "the sensorimotor system passing predictions of proprioceptive input to the classic reflex arcs, which fulfill them and thereby cause action" (Hohwy, 2016, p. 262). I reject this view and will develop an account on which active inference is construed as direct (not inferentially mediated) engagement with the environment (Bruineberg et al., 2018; Kiverstein, 2018). On this view, perceptual and active inference are intricately linked rather than one being in the service of the other. Active inference captures the action-oriented nature of PP which enables predictive control and has the positive effect of enabling an agent to act in order to regulate vital parameters. Importantly, it is the aim of the system to exercise predictive control by deciding which strategy to use for successful prediction error minimization in the long run. If the system always adapts to signals regardless of how noisy and uncertain they are, it runs the risk of overfitting the generative models—making it unreliable as a way to structure the world. On the other hand, not adapting the models when prediction error is propagated upwards, runs the risk of underfitting the model. The need to explore the environment and seek sensory information then becomes redundant. It is therefore important for the system to strike a delicate balance between changing the model and its parameters, on the one hand, and maintaining the parameters and changing the incoming signals.

The features discussed in this section form the foundation for an understanding of PP in terms of prediction error minimization as it is derived from the free energy principle. These features are interpreted in various ways and are highlighted to various degrees. I discuss two interpretations of predictive processing before developing the EEPP account. The first interpretation, I call cognitivist PP. This account is spearheaded by Jakob Hohwy

who refers to his account as prediction error minimization; it is also referred to as "conservative predictive processing" by Clark (2015). I refrain from using Hohwy's terminology to avoid confusion given that my own account makes use of prediction error minimization as a function but does not restrict this function to the neural domain. On Clark's terminology, my account would also be understood as conservative given that I do not propose to discard the notion of representation. But I grant a constitutive role to the body so I set my account apart from an internalist, cognitivist interpretation of PP. The second interpretation that I discuss arises from a combination of radically enactive cognition (REC) and "radical predictive processing" Clark (2015). I call this free energy enactivism to highlight the amplified role of the free energy principle in cognitive processing.

COGNITIVIST PREDICTIVE PROCESSING

The cognitivist interpretation of predictive processing builds on the features discussed above and construes the brain as a prediction error minimization system. Prediction errors signal the mismatches between bottom-up sensory signals and multi-area, top-down flows of *neuronal* activity (Clark, 2017b, p. 727, my emphasis) which serve to reconstruct the external reality. This process requires that the mind is an independent system that processes information entering from the outside world, reconstructing and mirroring the world for the agent to interact with. Anything that requires us to interact with it must be modeled. This distinction between mind and world enforces a strong and rigid evidentiary boundary between what happens in the external world and the generative models in the brain. In this sense, cognitive processes are inferentially secluded and neurocentrically skull bound (Hohwy, 2016, p. 259). Thus, any inputs beyond the sensory organs are outside the evidentiary boundary and can only be reconstructed (represented) in the brain. Hohwy (2016) epistemically decouples the brain from the body and world by suggesting that the brain, in implementing prediction error minimization, is self-evidencing. The brain has a model of the environment in which it is found and is continually updating generative models or changing input. It is equipped with the task of explaining away sensory input and, in doing so, it generates evidence for its own existence. This does not depict the brain as a passive organ; instead the brain is actively sampling evidence that matches its predictions and exploits the body as a tool in this undertaking. Perception is a process of representation only realized in the brain that infers distal information based on "partial and fragmentary information available in the sensory signal" (Clark, 2017b, p. 729). Our access to the world is bounded by prediction error minimization.

On this approach, action is explained in terms of proprioceptive prediction in that the approach construes action as a result of the brain's predictions about what state the body should be in (Friston and Stephan (2007), Friston (2010), and Hohwy (2016)). Action is an inferential process that starts in the neural domain and then "the body as it were goes away and does its own thing until the predictions come true" (Hohwy, 2016, p. 276). On the cognitivist PP approach, having embodied access to the world is not a necessary condition of the prediction error minimization system—it just so happens that we have bodies and therefore action is more likely (Hohwy, 2018, p. 135). Thus, predictive control is not explained in terms of agentive access to the world, or coupling between agent and environment, but rather in terms of the brain selectively sampling the sensory evidence presented to it (Burr and Jones, 2016). The brain is in the spotlight and the body in itself plays no constitutive role because "the mind begins where sensory input is delivered through exteroceptive, proprioceptive, and interoceptive receptors and it ends where proprioceptive predictions are delivered, mainly in the spinal cord" (Hohwy, 2016, p. 276). On this view, the body is important only insofar as it is represented in the neural hierarchy. Neural populations transmit commands for action based on sensory input. There is no direct access and engagement with the real world.

A notable implication of cognitivist PP is that the mind can be explained in entirely "internalist, solipsistic terms, throwing away the body, the world, and other people" (Hohwy, 2016, p. 265). The scope of cognitivist PP is thus limited to the brain, and all other phenomena (including the body and tools in the environment) only serve as resources to fulfill the function of prediction error minimization. Prediction error can be minimized using two strategies: (1) changing sensory input through action or (2) changing the internal models of the world. On the cognitivist PP account, both these strategies are explained as occuring primarily within the bounds of the skull. All processes relating to the agent are cashed out in terms of what happens in the cortical hierarchy. Action is enslaved in service of the brain and parts of our own bodies that are not functionally sensory organs are not constituents of cognitive states (Hohwy, 2016, p. 269). Bodily movements, as well as processes such as heart rate, are all inferred processes, lying beyond the evidentiary boundary. Construing the body as just another cause in the environment implies that it is nothing special, and neither is representation thereof (Hohwy and Michael, 2017). Although bodily movement is understood as facilitating prediction error minimization, and thus still a key feature in the cognitivist PP account, the role of the body is largely underplayed. Bodily movement is understood as an inferential process that arises from reconstructing the world rather than as enabled by sensory co-ordination.

Cognitivist PP does not grant the body any constitutive role in cognition. This is a symptom of the account taking a functionalist approach to explanation and limiting the function of prediction error minimization to the brain. The primary function of the brain, on this approach, is to minimize prediction error and all other phenomena serve only as tools to fulfill this function and are not explanatorily valuable in themselves. Hohwy (2015) sees value in a functionalist explanation because, he proposes, it provides a unifying principle for understanding what the brain does. Perception, for example, is specified in terms of a particular function—generating the best possible model of what is observed—then broken down into further sub-capacities such as estimating precision and fitting statistical models. These sub-capacities are then organized in a way that realizes the overall function of the capacity to be explained. Consider a

non-biological example of functional explanation: assembly line production (Cummins, 2000). In an assembly line, workers are assigned a task and the final product is successfully produced because each station has fulfilled their assigned function. The entire system can successfully fulfill its overall, unified function (producing a product) because each station fulfilled its given tasks in an organized way. An assembly line can be explained without making reference to the product being produced, the factory in which it is produced, and the number of stations involved in production. Similarly, a functional analysis limiting prediction error minimization to the brain does not make reference to the whole system that realizes the function but only to the function itself. Hohwy (2015, p. 17) acknowledges the problem of realization, and that a system has certain kinds of mechanisms that realize the function but limits talk of realization to neuronal circuitry. This approach is paradigmatic pure functionalism which is strongly committed to explaining only the functional role of a phenomenon and not how it is realized (Cummins, 2000; Egan, 2018). Although this can provide much insight into why the brain processes information in the way it does, and why we interact with the world in particular ways, the account leaves much desired in terms of explaining how prediction error minimization is realized. Providing an account of the "how" would require consideration of all components of the system including, I argue, the constitutive role of the body. In the next section, I discuss free energy enactivism which grants the body a central role in its explanations but at the cost of blurring the boundaries between what is understood as being cognitive and what is not.

FREE ENERGY ENACTIVISM

The fundamentally active and world-involving nature of predictive processing (PP) offers a point of agreement with enactivism. But despite the central role of action for cognition in PP, a tension arises because the PP framework does not seem to be complete without appeal to generative models that require contentful representations. Radically enactive cognition (REC) suggests that basic (i.e., not mediated by or involving language) cognition is contentless and non-representational (Hutto and Myin, 2013, 2017) and since PP is grounded in the manipulation of representational contents, the two accounts are in tension. REC outright rejects the cognitivist interpretation of PP and even a more "radical" version of PP that posits action-oriented representations. REC's objection is that any account that appeals to representations in its explanation must deal with the hard problem of content which involves explaining where the brain gets its conceptual resources from to represent information and make inferences (Hutto, 2018). According to REC, no acceptable answer has been offered by proponents of PP. Hutto (2018, p. 21) suggests that prediction error minimization can be explained in terms of embodied anticipations that are "grounded in structural and functional neural and other changes wrought through an organism's history of interactions." This implies that our actions and experiences change our neural setup not in terms of neural representations but rather in that the neural domain is "set up to be set off" (Prinz, 2004, p. 55). Thus, information processing is not the same as energy transfer or electrical activity in the brain but rather information-as-covariance (Hutto, 2018, p. 22). But the account offered by REC leaves much to be desired in that it does not provide a positive proposal about how else we could cash out the idea that the predictive system harbors generative models, that something or other is expected or predicted, and that there are matches or mismatches between top-down predictions and bottom-up signals.

Building on the same foundations as radical enactivism, another radical interpretation of PP has been developed in the literature; I call this interpretation free energy enactivism. Free energy enactivism, unlike cognitivist PP, proposes that the free energy principle and the inferential account of perception and cognition are conceptually independent (Bruineberg et al., 2018). The free energy enactivist approach maintains that the dynamic coupling between organism and world suffices to explain cognition and thus the notion of inference in the brain is not required. The premise for the free energy principle providing an account of cognition is that free energy is a function of sensory states and the internal dynamics of a biological system. This function is extended to the whole embodied organism, and not limited to reconstructing the structure of the environment in terms of representations. Instead, it is self-maintaining processes that endow an agent with a lived perspective and any disequilibrium shapes the way in which the world is perceived (Bruineberg et al., 2018, p. 2,426). Perception, on this view, is a result of the agent being open and responsive to affordances based on its metabolic and thermal disequilibria. Free energy enactivism thus understands perception as worthless without reference to action.

Free energy enactivism aims to provide an account that unifies biology and cognitive science. One of the radical claims put forward by free energy enactivism is that "the free energy principle applies not just to humans but to all living systems, including the simplest of life forms such as bacteria" (Bruineberg et al., 2018, p. 2,419). The principle has also been applied to plant cognition suggesting that plants predict the environmental factors that cause sensory stimulation (Calvo and Friston, 2017). Rather than appealing to the notion of representation, the generative models that predict the structure of the world has the function of mediating the organism's interactions with the world rather than reconstructing them. How does free energy enactivism appeal to models of the world without the notion of representation? Friston (2011; 2013b) suggests that it is not the case that an agent merely reconstructs a model of the world, but the agent is a model: the organism *embodies* an optimal model of its environment. In this sense, environmental features play a constitutive role in cognition; the internal and external morphology of an agent is constrained by the environment in which it is found. This is a bidirectional process because an organism's morphology also determines the environment in which the organism can survive. The interplay varies along timescales in that the agent may adapt to the environment in the long term but will change the environment for shorter term survival and efficiency.

Construed in this way, free energy enactivism illustrates a deep continuity between mind and life which is typical of enactive

approaches to cognition. On this view, the free energy principle applies to bacteria and plants as much as it applies to human agents in that these living systems engage in adaptive behavior (Kirchhoff and Froese, 2017). There is an implication that follows from this. If minimizing free energy is sufficient for mind and life, then all systems that resist disorder (or stay within bounds) exhibit mentality and are alive. There are two ways such a claim can be supported. Either one holds the view that mentality is not limited to living systems or by maintaining that life and mind are ubiquitous features (Kirchhoff and Froese, 2017). Both options give rise to panpsychism unless something further is added to the equation. The worry is that the scope of free energy enactivism is too broad in application and seemingly applies to non-living, non-cognitive systems. In other words, the boundaries between living, cognitive systems and the external non-cognitive world are blurred.

Furthermore, the free energy principle is construed as a nomological principle that all living systems abide by. It has been described as "normative" (Friston, 2013a), an "overarching rationale" (Clark, 2013), and a "law-like regularity" (Hohwy, 2013). On the radical construal by free energy enactivism, an organism is dynamically coupled with the environment through generalized synchrony (Friston, 2013b). But the notion of generalized synchrony is observed even in pendulum clocks that eventually synchronize through the beams from which they are suspended. This implies that one clock infers the state of another and is a generative model of the dynamics of the environment (Bruineberg et al., 2018, p. 2,437). Taking the nomological explanation presumed by free energy enactivism seriously means that all instances of generalized synchrony are instances of free energy minimization. And free energy minimization is a sufficient condition for a system to be a living system. The implication is that by applying a general law such as the free energy principle to dynamical systems, from pendulum clocks to human cognition, the explanatory value of the principle is lost. "Laws simply tell us what happens; they do not tell us why or how" (Cummins, 2000, p. 119). Arguably, the free energy principle can explain the capacities of dynamical systems but it does not follow that it can predict all capacities of such systems (or similar systems)—despite its ambitions to do so. The free energy principle serves well to explain the organization of dynamical systems, but it does not follow that the principle then adequately explains cognition. The free energy principle is very wide in scope and overshoots by trying to fully explain cognition. Rather than ambitiously attempting to explain all phenomena with a single principle, the aim should be to search for an explanation that captures the regularities of whole embodied organisms and their interaction with the environment. The free energy principle is presented as doing exactly this but I argue that despite how it is presented, the explanandum of the free energy principle under free energy enactivism is not the same as that of PP.

One way to sidestep the challenges is to consider the differences between non-living dynamical systems, simple life-forms and complex human agents where the latter may employ representational knowledge structures. But free energy enactivism rejects this position and suggests that an appeal to representation is not necessary. My proposal is that only by explaining additional components of the sophisticated system do we get an explanatorily useful account of perception, action and cognition. The free energy enactivist interpretation of PP also leaves much to be desired in terms of accounting for all the components of the system that realize prediction error minimization. I address this gap in the rest of this paper.

FINDING A THIRD WAY

Predictive processing (PP) is committed to providing causal and constitutive explanations of cognitive capacities. Achieving this requires investigating what kind of (methodological) explanation fits well with PP. I suggest that the explanatory methods of PP should be aligned with the mechanistic approach to explanation and that this requires including all components that realize cognition (including the body). Currently, both cognitivist PP and free energy enactivism offer no more than mere description and functional analysis, and though these accounts do not reject a mechanistic approach, they fail to include *all* components in their respective explanations. The two accounts that I have unpacked also differ in what they take to constitute cognition. Cognitivist PP restricts cognition to the neural organ and anything beyond that, including the body, only serves as a tool to minimize prediction error. The body as mechanism is reduced to how it contributes to prediction error minimization which is realized only in the neural domain. Free energy enactivism, on the other hand, extends cognition beyond the organism into the world such that the boundaries between cognitive and non-cognitive phenomena become blurred. I propose that these views represent extremes that alone do not successfully explain cognition. Instead, I defend a strongly embodied view that embeds the agent in the environment in which it is found (Friston, 2011; Pezzulo, 2014; Clark, 2015). I will argue that this is aligned with the mechanistic approach of explanation which identifies all relevant components of a system in realizing the phenomenon to be explained thus respecting both functional and structural properties. Jakob Hohwy, the key proponent of cognitivist PP, identifies the mechanistic potential of the framework but remains committed to a functionalist explanation of cognitive capacities in virtue of concepts such as precision, prediction error and model optimization (Harkness, 2015, p. 6). I suggest that PP can explain common sets of sub-capacities of cognition and their organization. This can then be used to provide an account of how cognitive capacities are realized in different biological systems. On the strong embodied view, the body is a constituent of cognition, i.e., it is part of the mechanisms that realize the function of the system. An account of PP should include an explanation that includes the body as realizer of prediction error minimization given that all components of the system and their capacities must be explained on the mechanistic view.

Explanation in cognitivist PP and free energy enactivism is aimed at describing free energy minimization (or prediction error minimization), but both accounts neglect to consider the structures and mechanisms that realize this phenomenon. This

is not to say that the contributions of these accounts are in vain but functional explanations can be enriched with mechanistic explanations (Piccinini and Craver, 2011; Harkness, 2015). Functional explanations often serve as first steps in mechanistic explanation in the sense that functional explanations provide sketches of mechanisms and the gaps are then later filled out (Piccinini and Craver, 2011, p. 284). Mechanistic explanations identify the relevant components of the system and respects the importance of both functional and structural properties. Given the importance of the structure of the system in which a phenomenon is realized, I suggest that mechanistic explanation serves PP better. Adopting a mechanistic approach enables the explanation of the capacities of the system and its component parts as opposed to only explaining the functions and effects of the system. The structures and processes that realize prediction error minimization are explained rather than merely describing it via functional analyses or nomological principles.

Mechanistic Explanation and Predictive Processing

Mechanistic explanation involves identifying the relevant parts of the mechanism, determining the operation they perform, and providing an account of how parts and operations are organized such that, under specific contextual conditions, the mechanism realizes the phenomenon of interest (Bechtel, 2009, p. 553). The Watt governor can serve as an example here and is often used as an analogy in dynamical systems theory of cognition. A Watt governor has the function of regulating the speed of engines. It functions to keep a system within a particular state (or range of states) and is constituted by several independent parts: the flywheel, the spindle and arms, and a type of linkage system connected to a valve. Each component of the governor operates on its own principles and performs a specific operation which contributes to the overall function of the system. It is because the spindle arms fulfill their function of rising and falling in response to the speed of the flywheel that their angle can be used to manipulate the linkage system. The spindle arms then open or close the valve allowing more or less fuel to pass through, increasing or decreasing the speed of the engine. The valve has no access to the speed of the flywheel without the spindle arms and linkage mechanism. All these mechanism form part of the system because they "encode" information that can be used by the valve. The Watt governor is a control system which is dependent on feedback to revise and redirect the behavior of parts of the mechanism.

There are uncanny similarities between the Watt governor as a mechanism that keeps the speed of engines within a particular range of states and the prediction error minimization system that functions to keep a living organism within a particular range of states. I suggest that the cognitive system comprises the whole embodied agent which includes the nervous system, the body, and relevant aspects of the environment. Like the Watt governor, each component of the embodied agent is a mechanism which operates on its own principles and performs specific operations, together contributing to the overall function of minimizing prediction error and keeping the agent within a particular range of states. The whole embodied agent is a control system and relies on feedback to control and direct motor activity and behavior. I suggest that prediction error minimization is not *only* performed through an interplay between predictions in the brain and activity at the sensory boundary (as proposed by cognitivist PP). Instead, we should think of prediction error minimization as the result of each component of the system (including the body) operating on its own principles and performing its own functions. For example, the body, in virtue of being a model of the environment, minimizes free energy by adapting accordingly across a long-term timescale. Prediction error minimization is realized by generative models in the brain and together with bodily movements the function is fulfilled. Representational mental states constitute only one component of the overall mechanistic system. The body is also a constitutive component in the process of minimizing prediction error and should not be treated as only a tool to fulfill the function of the brain. Each of the components of the system fulfills its own operations allowing the system to use the information to minimize prediction error in the long run.

EMBODIED, EMBEDDED PREDICTIVE PROCESSING

As I have unpacked in earlier sections, prediction error minimization and the free energy principle are not identical concepts. They differ in scope and explanation. This view has recently been argued for by Hohwy (2020) who proposes that the free energy minimization account provides a conceptual and mathematical analysis that is primarily a nomological explanation and PP offers a falsifiable process-theory that is a mere application of the free energy principle. Another approach that separates free energy minimization and prediction error minimization is offered by Bruineberg et al. (2018) who propose that perceptual inference is not compatible with the claims made by the free energy principle. Analyzing these two arguments lies beyond the scope of this paper but is worth mentioning as key players in the debate separating the two concepts. The account that I develop is based on the separation of prediction error minimization and the free energy principle. Yet it cannot be neatly separated from either, and it does not need to be because it does not reject the compatibility of the two concepts. EEPP fits into the larger ambitions of the free energy principle and is also a way of explaining how prediction error minimization is realized. In this sense, the account that I develop is more sympathetic to that of Hohwy (2020) as opposed to that of Bruineberg et al. (2018) because my account does not commit to the idea that free energy minimization and perceptual inference are incompatible. Rather, these processes are realized at different levels of the cognitive system and the different mechanisms of the system operate on their own principles. Making space for both these concepts in a single account provides at least one good reason to consider EEPP as a viable alternative approach. Prediction error minimization is a process that gives rise to perception and (to a certain degree)

enables action. Free energy minimization is implemented at the level of the whole embodied organism in virtue of the agent being a model of the environment. The embodied agent is embedded in the environment and engages in active inference to minimize uncertainty and disorder in the long term. Some insight about cognition can be derived *a priori* from the free energy principle, but the principle alone is too wide in scope to tell us all we want to know about cognition. Cognitivist PP, on the other hand, makes use of the free energy principle to develop an account of prediction error minimization but consequently restricts the scope of explanation to the neural domain underplaying the role of the body. The account that I develop will show that both these approaches contribute useful insights to our understanding of cognition but that by continuing to develop in opposing directions, the debate is losing sight of the phenomena in question: cognitive, embodied agents embedded in the world.

As cognitive, embodied agents we are directed at the world in a structured way. This capacity and our ability to act on the world is what sets us apart from other non-living systems. As I will argue in the next sections, the mechanisms that constitute cognition are not restricted to the neural domain. Prediction error minimization is instantiated not only by the neural domain but involves the whole system comprising the nervous system, body and relevant aspects of the environment (Anderson, 2014; Pezzulo, 2014; Clark, 2017a). I propose that prediction error minimization is deeply dependent on the body of an agent, such that the body-beyond-brain plays a constitutive role in cognitive processing. The body plays three constitutive roles in cognition:

1. The body *regulates* cognitive activity, ensuring that cognition and action are intricately linked. A prime example of this is the outfielder's problem.
2. The body acts as *distributor* in the sense that it carries some of the cognitive load of neural structures. This is illustrated by examples such as interoception and the use of gestures.
3. The body serves to *constrain* the information that is processed by an agent. This is supported by the idea that the agent is a model of the environment.

The descriptions of these roles are not separable in a very clear way and often a single example can be used to explain multiple roles. I unpack each of these roles in the following sections.

The Body as Regulator

The idea that cognitive processes serve to accommodate interaction with the world as opposed to reconstructing the world fits well to our understanding of the body as regulator. In embodied cognition approaches, the body as regulator thesis states that "an agent's body functions to regulate cognitive activity over space and time, ensuring that cognition and action are tightly coordinated" (Wilson and Foglia, 2017). The embodied PP account explains how agents are geared toward fast, successful, and fluent engagement with the environment, using simple routines and minimal representation. The whole embodied agent includes a cognitive system that is made up of several mechanisms each operating on its own principles of operation. The body serves as regulator insofar as it enables the agent to perceive and interact with the world through embodying the causal structure of the dynamics of the environment and itself. Successful movement and action in the world are possible because of coupling between agent and environment and does not necessarily require reconstructing the sensory signals. Consider the outfielder's problem: this scenario would involve a series of complex, action-sensitive information streams being fed to the brain—as if the agent is actually running to cancel the optical accelerations of the ball (Clark, 2017b, p. 735). The complexity involved in such a process would seem to count in favor of an account that can explain action and inference in simpler, embodied terms. This captures the notion of ecological efficiency which calls for a division of labor between brain, body, and environment. Division of labor between mind, body, and world enables the "productively lazy brain to do as little as possible while solving (or rather, while the whole, environmentally-located system) solves the problem" (Clark, 2015, p. 12). The cognitivist PP account can deliver an explanation of the outfielder's problem but not without "throwing away" the world and the body. For the cognitivist, the action-perception process involved in the outfielder's problem is one of inference that is a result of generative models that reconstruct a mirror of the world. The function of the system, on this account, is to generate hypotheses and find the best explanation of the sensory perturbations (the ball is moving and will drop to point x so in order to minimize prediction error, the outfielder must predict where the ball will land and then act in the world to move to point x). But on the embodied account that I develop, the function of the predictive system is to accommodate sensory perturbations to enable action in the world (the outfielder moves their body in such a way as to stay in a particular angle to the ball until meeting at the same point).

The embodied system is efficient because it uses minimal resources to capture what is necessary to act in the world. Navigating my way through a busy street is a complex task that requires movement of the body, adapting to uneven sidewalks, avoiding running children and other obstacles. The body regulates the agent's interaction with the environment in virtue of the coupled dynamics between the environment and the body. This means commanding models with the least prediction error or with the least sensory signal to "explain away." This notion requires an evidentiary boundary to distinguish between inferences and what is predicted. Cognitivist PP takes this boundary to be solid and clear ". . .with the brain on one side and the worldly and bodily causes on the other side" (Hohwy, 2016, p. 281). But on the embodied approach, the boundary becomes flexible and immutable (Clark, 2017c; Kirchhoff and Kiverstein, 2019). This does not mean the boundary does not exist—this would lead to the dissolution of the predictive task[2].

[2]The debate on the nature and how far out the boundary extends beyond the neural domain is still hotly debated. Kirchhoff and Kiverstein (2019) defend an extended mind view and propose that the boundary (demarcated by the Markov blanket) extends all the way out into the world. On my account, the boundary is determined by the lived body of the agent. Concretely, this means beyond the neural domain to include the body but not including tools and other resources out in the world.

Instead, the boundary is determined by the agent and her *lived* body. It is not necessary for the body of the agent to be modeled and predicted in the same way as the external world because it does not lie outside the boundary. The boundary is determined by the physical lived body of the agent insofar as the agent embodies the causal structure of the environment which gives rise to a state of action readiness; the embodied agent is ready to act on the salient action possibilities in the environment. As active systems, we are constantly seeking which sensory input to sample next instead of passively matching prior probabilities with states of the environment. The body is crucial to the successful execution of this task because without it, there would be no interaction in the world, nor would there be any prediction error to minimize. The embodied PP account claims that the brain minimizes prediction error to accommodate the sensory barrage. Accommodating the sensory barrage involves other low-cost methods that do not imply action-neutral modeling of the environment.

The Body as Distributor

The explanation above fits well with another way in which the body plays a constitutive role in cognition: as distributor. The body as distributor thesis states that "an agent's body functions to distribute computational and representational load between neural and non-neural structures" (Wilson and Foglia, 2017). In the PP account, this means that prediction error is minimized by both neural and non-neural structures, such as the body-beyond-brain. A similar view is also put forward by Bruineberg et al. (2018) who propose that the predictive neural system does not "know" about the viable states in which the agent must maintain its body (a certain temperature, for example) and therefore such an embodied state can only be maintained by the body itself, i.e., without neural mediation. They call this embodied surprisal and use it as a premise to argue for the incompatibility of the free energy principle and prediction error minimization. Although I agree that the body can realize the function of prediction error minimization without neural mediation, I do not propose that these processes are separate and incompatible but rather that prediction error minimization in the neural domain and the minimization of, so-called, embodied surprisal are intricately linked.

On my account, action can be described as a process of inference that uses a non-reconstructive strategy to keep certain sensory stimulations within bounds. It is thus not necessary to reconstruct a model of the real world to plan, reason and guide successful behavior and action. Instead interaction with the environment is "a kind ofperceptually-maintained motor-informational grip on the world: a low-cost perception-action routine that retrieves the right information just-in-time for use" (Clark, 2017b, p. 737). The idea of body as distributor can be explored in EEPP by looking at how interoceptive information is processed. Perception of the body plays an important role in how we represent the world. For example, imagine you are watching a horror movie. As a result, your attention increases and your heartbeat accelerates. You hear a sound just outside the window which can be caused by several things. For the purpose of this example, let us limit the pool of hypotheses to two: (1) the wind is blowing a tree branch against the window, or (2) a thief is trying to gain entry into your house. Let us suppose you live in a low-crime area and have never experienced a break-in. The hypothesis with the highest prior probability should be that the wind is blowing a tree branch against the window. But given the interoceptive information and physiological state of your body, the thief-hypothesis has higher prior-probability[3]. This is because all the evidence (including interoceptive information) has to be explained. All available sensory information makes up the evidence against which a hypothesis is tested. Importantly, this sensory information is not limited to seeing hearing, smelling, tasting, and touch but also includes kinesthetic, proprioceptive and interoceptive information. In order to most effectively reduce prediction error, the whole embodied agent is involved. The body of the agent (in the above case, through interoception) contributes to the minimization of prediction error because it carries useful and reliable information.

The Body as Constraint

The body as constraint thesis states that: "an agent's body functions to significantly constrain the nature and content of representations processed by that agent's cognitive system" (Wilson and Foglia, 2017). On PP, this can be understood in terms of how the agent models the environment. There are two ways in which the embodied agent models the environment. First, in terms of embodying a model of the environment, i.e., being a model of the environment. Second, in terms of generating action-oriented models of the world, i.e., having a model of the environment. Explaining in detail the two ways in which an agent models the environment in virtue of being an embodied agent requires more space than the scope of this paper allows and so the exposition that follows is brief. First, the embodied agent is not only modeled in the predictive system as part of the outside world but also acts as the point of reference from which the world is perceived. Interaction with the environment is made possible not only because the agent generates models of the world but the agent is its own best possible model of the world (Bruineberg et al., 2018, p. 2,425). The agent embodies a model of the environment in virtue of the coupled relation of internal and external dynamics, i.e., the structure of the environment is reflected in the embodied agent. In this sense, the environment and the embodied agent structure and constrain one another (Bruineberg et al., 2018, p. 2,422).

Second, the models that are generated in the predictive system are constrained by the structure of the embodied agent. Representing the world involves representing properties of objects such as shape, color, size and location but possibilities for action are also modeled and these affordances are only modeled as they are relevant and salient to the embodied agent. The affordance of sitting on a chair is only available to me, a human agent, insofar as I have the necessary limbs and joints that make this possible. My body thus constrains the models of the world that are generated; if I am paraplegic, a chair does not afford sitting but is rather an obstacle that I must

[3]One could argue that another way of thinking about this is in terms of cognitive bias, for example if you were primed into expecting a thief because this a break-in occurred in the film you were watching. But this explanation does not suffice because interoceptive information is often more reliable than sensory stimuli.

avoid while moving around in my wheelchair. Most compatible with the embodied PP account is the notion of action-oriented representations. Action-oriented representations are aimed at driving specific action and are not reconstructive and detached from the world, nor are they disembodied (i.e., independent from the agent and their abilities). Action-oriented representations encode the affordances of objects as they are relevant and salient to the agent. Part of the predictive task is to anticipate and discriminate between things in the environment that matter to an agent and those that do not. In this sense, "the brain is constantly computing—partially and in parallel—a large set of possible actions" (Clark, 2016, p. 180). Concretely, this implies that the generative models in the predictive system are not detached and neutral reconstructions of the world but rather generative models of the possible ways in which the agent can interact with the world as constrained by their body. Such action-oriented generative models enable fluent interaction with the world because they are generated from the perspective of the agent, i.e., it is specifically relevant and salient to them in virtue of being an embodied agent, embedded in a specific environment.

CONCLUSION

In this paper, I distinguished between two radical interpretations of the predictive processing framework. The divergence between the two positions is motivated by the conceptual distinction between the free energy principle and inferential perception (realized as prediction error minimization). As an alternative position, I propose a strongly embodied interpretation of predictive processing that take the whole embodied agent as well as relevant aspects of the environment to realize prediction error minimization. This alternative position includes the body as a constitutive part of cognition and as realizer of prediction error minimization. It also includes relevant aspects of the environment to constitute prediction error minimization. This can be understood in terms of affordances. Rather than include the whole environmental system in cognitive function (as proposed by free energy enactivism), I propose that only the brain and body-beyond-brain form part of the cognitive system. This implies that the boundary between cognitive and non-cognitive phenomena is not rigid and predetermined but rather flexible and immutable. Developing a full account of EEPP is an enormous undertaking and requires contributions from many fields of science and philosophy. This paper aimed to deliver a starting point for such developments in the field rather than develop a fully fleshed out account.

[4] There is a general divide between the action-first approach—construing affordances as byproducts of action plans—and the spectator-first approach—which highlights the role of belief-like representations of scenes with which an agent does not necessarily interact (Siegel, 2014, p. 51). I defend the view that affordances are relations between aspects of the environment and the abilities of an agent. This is in line with free energy enactivism which also maintains that affordances stand out as relevant in a specific situation lived by the agent and constitute the (pre-reflective) experiential equivalent of bodily action readiness: "the readiness of the affordance-related ability" (Bruineberg and Rietveld, 2014, p. 2).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - Projektnummer GRK-2185/1 (DFG-Graduiertenkolleg Situated Cognition). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number GRK-2185/1 (DFG Research Training Group Situated Cognition).

ACKNOWLEDGMENTS

I am grateful to Prof. Dr. Tobias Schlicht and Dr. Krzysztof Dołęga for their very helpful comments on this research. Thank you to the two reviewers for their very helpful and constructive feedback.

REFERENCES

Alsmith, A. J. T., and de Vignemont, F. (2012). Embodying the mind and representing the body. *Rev. Philos. Psychol.* 3, 1–13. doi: 10.1007/s13164-012-0085-4

Anderson, M. L. (2014). *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: MIT Press.

Bechtel, W. (2009). Constructing a philosophy of science of cognitive science. *Top. Cogn. Sci.* 1, 548–569. doi: 10.1111/j.1756-8765.2009.01039.x

Bruineberg, J., and Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Front. Hum. Neurosci.* 8:599. doi: 10.3389/fnhum.2014.00599

Bruineberg, J., Rietveld, E., and Kiverstein, J. (2018). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese* 195, 2417–2444. doi: 10.1007/s11229-016-1239-1

Burr, C., and Jones, M. (2016). The body as laboratory: prediction-error minimization, embodiment, and representation. *Philos. Psychol.* 29, 586–600. doi: 10.1080/09515089.2015.1135238

Calvo, P., and Friston, K. (2017). Predicting green: really radical (plant) predictive processing. *J. R. Soc. Interface* 14:20170096. doi: 10.1098/rsif.2017.0096

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Clark, A. (2015). Radical predictive processing. *South. J. Philos.* 53, 3–27. doi: 10.1111/sjp.12120

Clark, A. (2016). *Surfing Uncertainty*. New York, NY: Oxford University Press.

Clark, A. (2017a). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenol. Cogn. Sci.* 17, 521–534. doi: 10.1007/s11097-017-9525-z

Clark, A. (2017b). Busting out: predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Nous* 51, 727–753. doi: 10.1111/nous.12140

Clark, A. (2017c). "How to knit your own markov blanket?: resisting the second law with metamorphic minds," in *Philosophy and Predictive Coding*, eds T. Metzinger, and W. Wiese (Frankfurt am Main: MIND Group), 1–19. doi: 10.15502/9783958573031

Colombo, M., and Wright, C. (2016). Explanatory pluralism: an unrewarding prediction error for free energy theorists. *Brain Cogn.* 112, 3–12. doi: 10.1016/j.bandc.2016.02.003

Cummins, R. (2000). "How does it work?" versus "what are the laws?": Two conceptions of psychological explanation," in *Explanation and Cognition*, eds F. C. Keil, and R. A. Wilson (Cambridge, MA: MIT Press), 117–144.

Egan, F. (2018). "'Function-theoretic explanation and the search for neural mechanisms'," in *Explanation and Integration in Mind and Brain Science*, ed. D. M. Kaplan (Oxford: Oxford University Press), 145–163.

Fotopoulou, A., and Tsakiris, M. (2017). Mentalizing homeostasis: the social origins of interoceptive inference-replies to commentaries. *Neuropsychoanalysis* 19, 71–76. doi: 10.1080/15294145.2017.1307667

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K. (2011). "Embodied inference: or "i think therefore i am, if i am what i think"," in *The Implications of Embodiment (Cognition and Communication)*, eds W. Tschacher, and C. Bergomi (Exeter: Imprint Academic), 89–125. doi: 10.1007/978-3-319-75726-1_8

Friston, K. (2013a). Active inference and free energy. *Behav. Brain Sci.* 36, 212–213. doi: 10.1017/s0140525x12002142

Friston, K. (2013b). Life as we know it. *J. R. Soc. Interface* 10:20130475. doi: 10.1098/rsif.2013.0475

Friston, K., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y

Harkness, D. L. (2015). "From explanatory ambition to explanatory power," in *Open MIND*, eds T. Metzinger, and J. Windt (Frankfurt am Main: MIND Group), doi: 10.15502/9783958570153

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.

Hohwy, J. (2015). "The neural organ explains the mind," in *Open MIND*, eds T. Metzinger, and J. Windt (Frankfurt am Main: MIND Group), doi: 10.15502/9783958570016

Hohwy, J. (2016). The self-evidencing brain. *Nous* 50, 259–285. doi: 10.1111/nous.12062

Hohwy, J. (2018). "'Predictive processing hypothesis'," in *The Oxford Handbook of 4E Cognition*, 1st Edn, eds A. Newen, L. De Bruin, and S. Gallagher (Oxford: Oxford University Press), 127–146.

Hohwy, J. (2020). Self-supervision, normativity and the free energy principle. *Synthese*. 1–25. doi: 10.1007/s11229-020-02622-2

Hohwy, J., and Michael, J. (2017). "Why should any body have a self?," in *The Subject's Matter: Self-Consciousness and the Body*, eds F. de Vignemont, and A. Alsmith (Cambridge MA: MIT Press), 363–391.

Hutto, D. (2018). Getting into predictive processing's great guessing game: bootstrap heaven or hell?. *Synthese* 195, 2445–2458. doi: 10.1007/s11229-017-1385-0

Hutto, D., and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge MA: MIT Press.

Hutto, D., and Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content*. Cambridge MA: MIT Press.

Kirchhoff, M. (2017). Predictive brains and embodied, enactive cognition. *Synthese* 195, 2355–2366. doi: 10.1007/s11229-017-1534-5

Kirchhoff, M. (2018). "'The body in action: predictive processing and the embodiment thesis'," in *The Oxford Handbook of 4E Cognition*, eds A. Newen, L. de Bruin, and S. Gallagher (Oxford: Oxford University Press), 243–260.

Kirchhoff, M., and Froese, T. (2017). Where there is life there is mind: in support of a strong life-mind continuity thesis. *Entropy* 19, 1–18. doi: 10.3390/e19040169

Kirchhoff, M. D., and Kiverstein, J. (2019). How to determine the boundaries of the mind: a Markov blanket proposal. *Synthese*. 1–20. doi: 10.1007/s11229-019-02370-y

Kiverstein, J. (2018). Free energy and the self: an ecological–enactive interpretation. *Topoi* 39, 559–574. doi: 10.1007/s11245-018-9561-5

Metzinger, T., and Wiese, W. (2017). *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group.

Pezzulo, G. (2014). Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cogn. Affect. Behav. Neurosci.* 14, 902–911. doi: 10.3758/s13415-013-0227-x

Piccinini, G., and Craver, C. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese* 183, 283–311. doi: 10.1007/s11229-011-9898-4

Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. New York, NY: Oxford University Press.

Rowlands, M. (2010). *The New Science of the Mind*. Cambridge MA: MIT Press.

Rupert, R. D. (2009). *Cognitive Systems and the Extended Mind*. New York, NY: Oxford University Press.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* 17, 565–573. doi: 10.1016/j.tics.2013.09.007

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

Siegel, S. (2014). "'Affordances and the contents of perception'," in *Does Perception Have Content?*, ed. B. Brogaard (Oxford: Oxford University Press), 51–75. doi: 10.1093/acprof:oso/9780199756018.003.0003

Sims, A. (2016). A problem of scope for the free energy principle as a theory of cognition. *Philos. Psychol.* 29, 967–980. doi: 10.1080/09515089.2016.1200024

Wilson, R., and Foglia, L. (2017). *Embodied Cognition, The Stanford Encyclopedia of Philosophy*. Available online at: https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/ (accessed August 15, 2019).

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 04 February 2021
doi: 10.3389/fpsyg.2021.623970

The Network Theory of Psychiatric Disorders: A Critical Assessment of the Inclusion of Environmental Factors

Nina S. de Boer[1], Leon C. de Bruin[1,2], Jeroen J. G. Geurts[3] and Gerrit Glas[2,3]*

[1]Department of Philosophy, Radboud University, Nijmegen, Netherlands, [2]Department of Philosophy, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [3]Department of Anatomy and Neurosciences, Amsterdam University Medical Centers (Location VUmc), Amsterdam, Netherlands

Borsboom and colleagues have recently proposed a "network theory" of psychiatric disorders that conceptualizes psychiatric disorders as relatively stable networks of causally interacting symptoms. They have also claimed that the network theory should include non-symptom variables such as environmental factors. How are environmental factors incorporated in the network theory, and what kind of explanations of psychiatric disorders can such an "extended" network theory provide? The aim of this article is to critically examine what explanatory strategies the network theory that includes both symptoms and environmental factors can accommodate. We first analyze how proponents of the network theory conceptualize the relations between symptoms and between symptoms and environmental factors. Their claims suggest that the network theory could provide insight into the causal mechanisms underlying psychiatric disorders. We assess these claims in light of network analysis, Woodward's interventionist theory, and mechanistic explanation, and show that they can only be satisfied with additional assumptions and requirements. Then, we examine their claim that network characteristics may explain the dynamics of psychiatric disorders by means of a topological explanatory strategy. We argue that the network theory could accommodate topological explanations of symptom networks, but we also point out that this poses some difficulties. Finally, we suggest that a multilayer network account of psychiatric disorders might allow for the integration of symptoms and non-symptom factors related to psychiatric disorders and could accommodate both causal/mechanistic and topological explanations.

Keywords: network theory, network analysis, causality, interventionism, mechanistic explanation, topological explanation, multilayer network, psychiatry

OPEN ACCESS

Edited by:
Regina E. Fabry,
Ruhr University Bochum, Germany

Reviewed by:
Sam Wilkinson,
University of Exeter, United Kingdom
Matteo Colombo,
Tilburg University, Netherlands
Mark Daniel Miller,
University of Sussex, United Kingdom

****Correspondence:***
Nina S. de Boer
nina.deboer@ru.nl

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *30 October 2020*
Accepted: *18 January 2021*
Published: *04 February 2021*

Citation:
de Boer NS, de Bruin LC, Geurts JJG and Glas G (2021) The Network Theory of Psychiatric Disorders: A Critical Assessment of the Inclusion of Environmental Factors. Front. Psychol. 12:623970. doi: 10.3389/fpsyg.2021.623970

INTRODUCTION

How should we explain why and how symptoms of psychiatric disorders arise? According to a long-established view, this can be done by conceptualizing symptoms as the effects of a common cause. Proponents of this view (henceforth referred to as the *traditional view*) often assume that the common cause in question is neurobiological in nature, and thus (often implicitly) endorse the idea that psychiatric disorders can be explained in terms of lower-level, (neuro)biological properties. The influence of this view in the scientific debate is most convincingly exemplified by an article published by the former heads of the National Institute of Mental Health (NIMH) in *Science* titled "Brain disorders?, Precisely", stating that new diagnostics will likely redefine mental disorders as "brain circuit disorders" (Insel

and Cuthbert, 2015). Their claims are in line with the NIMH's Research Domain Criteria (RDoC) initiative, a now widely adopted framework that aims to transform our current diagnostic frameworks for psychiatric disorder classification into a biological system that "conceptualizes mental illnesses as brain disorders" (Insel et al., 2010, p. 749). Despite the influence of the traditional view, however, there is not much empirical evidence to support it. As Adam (2013 p. 417) puts it: "Despite decades of work, the genetic, metabolic, and cellular signatures of almost all mental syndromes remain largely a mystery." To illustrate, a recent meta-analysis on 73 potential biomarkers for obsessive-compulsive disorder demonstrated that none had sufficient sensitivity or specificity (Fullana et al., 2020).

A promising alternative account of psychiatric disorders that has gained traction over the past years is the *network theory*, which conceptualizes psychiatric disorders as relatively stable networks of interacting symptoms (Borsboom, 2017; Borsboom et al., 2019a).[1] Although network science has been around since the late twentieth century (Barabási, 2012), its application to psychopathology is fairly recent and provides a new way of understanding and explaining psychiatric disorders. Whereas proponents of the traditional view typically argue that the causes of psychiatric disorders are localizable in the brain, the network theory moves our focus from the brain to psychiatric symptoms and their relations. Proponents of the network theory (e.g., Borsboom, 2017; Borsboom et al., 2019a) have argued that the theory should not only focus on the symptom network, however, but should also include non-symptom factors relevant in the context of psychiatry, such as *environmental factors*. Examples are adverse life events, social relations, but also more pragmatic items such as external objects (e.g., gambling machines in gambling addiction, Borsboom et al., 2019a).[2] The underlying motivation is that different factors are involved in the development and sustenance of psychiatric disorders, and that we can only properly understand and explain these disorders if we take these factors and their relation to each other into account (Kendler, 2008; Nolen-Hoeksema and Watkins, 2011). How are environmental factors incorporated in the network theory, and what kind of explanations of psychiatric disorders can such an *extended* network theory provide? Addressing these questions is important because proponents of the network theory do not just want to use network models as instruments to investigate psychiatric disorders: they want to provide a theory of what psychiatric disorders *are* (Borsboom et al., 2019b). Although they have made various claims on the role of environmental factors in the network theory and the theory's explanatory potential, these claims would benefit from further justification.

The aim of this article is to critically examine what explanatory strategies the network theory that includes both symptoms and environmental factors can accommodate. First, we will analyze how proponents of the network theory conceptualize the relations between symptoms and between symptoms and environmental factors. We will focus primarily on the accounts of Borsboom (2017) and Borsboom et al. (2019a), since these are seminal papers on the network theory of psychiatric disorders and also make various claims on the causal and/or constitutive role of symptoms and environmental factors in relation to psychiatric disorders. Afterwards, we will examine if we can corroborate these claims using network analysis, Woodward's interventionist theory of causation, or mechanistic explanation. Next, we will examine the claim that the network theory can explain the dynamics of psychiatric disorders by referring to the network's characteristics by means of a topological explanatory strategy. Finally, we will introduce the *multilayer* network account of psychiatric disorders as a framework that allows for the integration of symptoms and non-symptom factors related to psychiatric disorders, and could potentially accommodate both causal/mechanistic and topological explanations.

THE NETWORK THEORY OF PSYCHIATRIC DISORDERS

The Symptom Network

Borsboom and colleagues make two main claims about relations in the symptom network. The first claim concerns the relations between symptoms. The network theory states that psychiatric symptoms *causally* interact with each other (Borsboom, 2017). This causal interpretation of the covariance between symptoms is justified by referring to folk psychology: they claim that it *makes sense* for certain symptoms to be causally related (Borsboom et al., 2019a).[3] It seems to make sense, for example,

[1]In this article, we distinguish between network *theory* and network *analysis*. We use the term network analysis to refer to the statistical techniques used to estimate networks based on empirical data. This term can be used synonymously with network methodology and network psychometrics. Network theory aims to address and explain the nature of psychopathology and to give an account of what psychiatric disorders *are* (Borsboom et al., 2019b). We will discuss this distinction more thoroughly in section "The causal/ mechanistic explanatory strategy."

[2]One of the anonymous peer reviewers alluded us to the article by Colombo and Heinz (2019) that assesses which theoretical framework can best integrate different aspects of psychiatric disorders. More specifically, they address how computational phenotypes and phenomenological information could be integrated into one explanatory account of alcohol use disorder. Similar to our article, Colombo and Heinz (2019) propose that such an integrative account should include multiple layers, and they discuss network models as one of the possibilities for explanatory integration. They argue that networks cannot include multiple layers (i.e., are *flat*), and claim that a dimensional model may be a more promising framework for explanatory integration. We agree that dimensional models may also be of interest, but it is important to note that the network theory (and the Borsboom and Cramer, 2013 article they make reference to) does not reject the possibility that a network may consist of multiple layers. This will be further addressed in section "A multilayer network account of psychiatric disorders."

[3]Borsboom et al. (2019a) claim that the relations between symptoms make sense by referring to interpretivism, i.e., the notion that we attribute beliefs, emotions, and desires with specific content to ourselves and others explain and predict behavior (Dennett, 1987). On their account, we can make sense of and understand why one symptom can lead to another by referring to their intentional content, i.e., what they are about, and people's *basic rationality*. For example, if one beliefs they may be spreading germs, it makes sense that they wash their hands excessively, since hand washing is a reasonable strategy to prevent the spreading of germs. Issues with this interpretation have been raised (e.g., Slors et al., 2019), but discussing this goes beyond the scope of this article.

that insomnia can lead to fatigue and that hallucinations can lead to the development of delusions (Kendler et al., 2011). However, those critical of the network theory could argue that intuition and sense-making are not necessarily reliable criteria for determining causality.

The second claim concerns the relation between symptoms and the psychiatric disorder in question. The network theory claims that the (causal) interactions between the symptoms themselves is *constitutive* of the disorder, rather than symptoms being caused by an underlying disorder.[4] To illustrate the difference between these views, consider the diagnostic criteria for major depressive disorder (MDD). According to the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders*, receiving a MDD diagnosis requires at least five of the following nine symptoms to be present almost every day during the same 2 week period: (1) depressed mood, (2) diminished interest or pleasure, (3) significant weight loss or gain, (4) insomnia or hypersomnia, (5) psychomotor agitation or retardation, (6) fatigue or loss of energy, (7) feelings of worthlessness or excessive/inappropriate guilt, (8) diminished ability to think/concentrate or indecisiveness, and (9) recurrent thoughts of death/suicidal ideation (American Psychological Association, 2013).[5] The traditional view would argue that MDD is the latent or unobserved cause of all these symptoms: to treat MDD, the disorder itself should be treated, after which the symptoms should also disappear. The network theory, however, would claim that MDD is constituted by the relatively stable configuration of causal interactions between the symptoms: to treat MDD, the symptoms should be treated directly. As argued by Borsboom (2017, p. 10): "If diagnosis involves identifying a symptom network, then treatment must involve changing or manipulating that network." But claiming that symptoms constitute a disorder also poses some issues. Since there is considerable variation in the type of symptom combinations one can have in order to receive an MDD diagnosis, how can we claim that these diverse combinations all constitute the same disorder?[6] This example illustrates that the network theory may benefit from justification criteria for their claims concerning causality and constitution in symptom networks.

The Role of Environmental Factors

Proponents of the network theory have also made various claims on the role of environmental factors in psychiatric disorders. First, it is sometimes claimed that symptoms and environmental factors are causally related, but that this causal relation is different from the causal relations between symptoms. Whereas it is considered that there may be feedback loops between individual symptoms, causal connections between symptoms and environmental factors are typically presented as *unidirectional*: environmental factors affect symptoms. Indeed, environmental factors are typically presented as catalysts or *background elements* of the symptom network: symptoms can be "activated by factors external to the person" (Borsboom et al., 2019a, p. 4), but the symptom network eventually becomes self-sustaining after activation. For example, losing one's partner may lead to a depressed mood, which can lead to insomnia, anxiety, etc. (Borsboom, 2017). It has also been claimed that environmental factors can influence and determine the strength of the relations between the symptoms (Borsboom et al., 2019a), hence directly influencing symptom-symptom relations.

The relation between environmental factors and symptoms is not only presented as causal, however. It is also claimed that environmental factors can be *constitutively* related to symptoms, to symptom-symptom relations, and to the disorder itself. This constitutive relation is presented by the claim that environmental factors can be part of the *mechanisms* that constitute the disorder: "(network structures) rest on or invoke mechanisms in the environment (Borsboom et al., 2019a, p. 8)." Concerning the constitutive role of environmental factors in symptom-symptom relations, proponents of the network theory claim that "we should expect to find interactions between symptoms to be grounded in an even more complex set of biological, social, and cultural factors involved in psychopathology" (Borsboom et al., 2019a, p. 10). To illustrate this, Borsboom and colleagues examine the role of a Roulette table in gambling addiction. They state that the relationship between excessive gambling and debt – both symptoms of gambling addiction – is realized by the gambling setups that require a monetary investment, for example, in the form of a Roulette table. If we imagine a world without Roulette tables, or with Roulette tables that are operationalized in a different way, there would not be a link between excessive gambling and debt. Hence, they claim that environmental factors (such as Roulette tables) are an integral part of the symptom-symptom relation. The network theory also argues that environmental factors can co-constitute a psychiatric disorder: "in network models (…) the environment itself may become part of the network structure, and hence part of the disorder. More or less by definition, this means that (…) cultural and historical factors as well as external mechanisms, to some extent, shape mental disorders" (Borsboom et al., 2019a, p. 8). Hence, whereas they argue that environmental factors can causally influence the symptom network, they also claim that environmental factors can be part of the disorder itself.

This demonstrates that proponents of network theory suggest various ways to interpret the relation between environmental factors and symptoms: environmental factors may cause or constitute symptoms and/or symptom-symptom connections, and may co-constitute the psychiatric disorder in question. Hence, these claims suggest that the network theory could explain the causal mechanisms underlying psychiatric disorders. Are these claims justified? How can we evaluate them? In the

[4]This claim is not made explicitly by Borsboom (2017) or Borsboom et al. (2019a), but it has been endorsed and explained in Borsboom (2008), Fried and Cramer (2017), and Oude Maatman (2020).

[5]The first two symptoms – depressed mood and diminished interest – are considered *core symptoms*, meaning that at least one of them needs to be present. Additionally, to receive a MDD diagnosis, the symptoms need to cause clinically significant distress and the episode should not be attributable to a substance or another medical condition or disorder.

[6]It should be noted that the notion of *disorder heterogeneity* also poses a problem for the traditional view: how can we justify referring to a common cause when there is substantial heterogeneity in the way psychiatric disorders are manifested? One possible albeit controversial means to solve this problem is to argue that different symptom manifestations constitute different disorders, but discussing this alternative in depth is beyond the scope of this article.

next section, we will assess these questions in relation to network analysis, Woodward's interventionist theory of causation and mechanistic explanation.

THE CAUSAL/MECHANISTIC EXPLANATORY STRATEGY

Network Analysis

It seems like a logical starting point to attempt to corroborate the aforementioned causal claims using statistical evidence since network theory has its origins in network analysis (Borsboom, 2008). Network analysis refers to statistical techniques that estimate (i.e., approximate "true, real-world") networks based on patterns of covariance in empirical data. These techniques generally estimate the relations between variables as *partial* correlations, i.e., associations between two traits conditioned on the other traits in the model.[7] A partial correlation between variables *A* and *B* in a network can be interpreted as the value of variable *A* predicting the value of variable *B*. For example, Beard et al. (2016) demonstrated a statistically significant relationship between depressed mood and diminished interest in a symptom network for individuals with a MDD diagnosis. This may indicate that mood changes in MDD predict changes in interest, and vice versa. Partial correlations between symptoms and environmental factors have been estimated in a similar fashion: studies have examined cannabis use, developmental trauma and urban environment (Isvoranu et al., 2016), sexual risk (Choi et al., 2017), and spousal loss (Fried et al., 2015) in relation to symptoms of a variety of psychiatric disorders. Some studies demonstrated that environmental factors may indeed predict symptoms (e.g., spousal loss is strongly associated with loneliness, Fried et al., 2015).

There are various reasons why we should not conflate the network theory with network analysis, however, as highlighted by Fried (2020) and Robinaugh et al. (2020). First, statistically estimating relations in a network is not a theory-neutral process: there are various choices that have to be made before one can claim that a relation is present or absent. For example, we can vary the threshold used for determining statistical significance and have to decide which regularization techniques to use to correct for false positives (Epskamp et al., 2017). Second, most statistical analyses – including all the aforementioned studies – use cross-sectional, between-subject data. Identifying a relation in a between-subject design does not necessarily provide information on whether this relation is present *within* a person (Fisher et al., 2018). Although within-subject network studies are being conducted (Bringmann et al., 2013), they still constitute the minority of the studies available. Third, the boundary between statistical network models and latent variable models is more nuanced than commonly assumed (Bringmann and Eronen, 2018). These models may be *statistically equivalent*: they may fit the same dataset equally well, meaning that they cannot provide enough evidence to promote one model over the other.

A final, important reason is that the network theory wants to do more than merely predict psychiatric disorders: it wants to provide *causal* explanations. If we know the causal processes underlying psychiatric disorders, we can come up with interventions and design suitable treatments or prevention programs accordingly. We cannot simply assume that (partial) correlations imply causal relations: covariance does not necessarily imply that one of the variables *influences* the other. As the classic example of the barometer and the storm goes: one can predict a storm using a barometer, but changing the pressure readings will not prevent the storm from happening. Relatedly, the presence of (partial) correlations does not rule out the traditional view that symptoms of psychiatric disorders have a common (brain-based) cause. Indeed, symptom covariance can still be explained under the traditional view that symptoms are caused by an underlying (neurobiological) cause. Now one could argue that causal inference techniques can be used to directly estimate *directed acyclic graphs* (DAGs), i.e., causal networks without bidirectional effects or feedback loops, using correlational data (Pearl, 2000). Indeed, DAGs have been used to study the causal relations between symptoms (Borsboom and Cramer, 2013), and between environmental factors and symptoms (Moffa et al., 2017). It is important to note, however, that these causal inference methods require certain assumptions to be satisfied. They assume that the network encodes all the causal relations between factors, that there is no unobserved confounding, and that there are no causal feedback loops.[8] These assumptions may not be met in the context of psychiatric disorders and will be discussed in more detail in the upcoming section.

Hence, we cannot corroborate the causal claims of the network theory based on network analysis alone: although statistical models can generate findings that need to be explained, they do not have the explanatory power that the theory claims to have.

Woodward's Interventionist Theory of Causation

Another potential means to justify the causal claims made by proponents of the network theory is to make reference to (hypothetical) interventions. This is also alluded to by proponents of the network theory: Borsboom (2017, p. 6) argues that "such causal interaction between symptoms can be interpreted using interventionist theories of causation." The interventionist theory of Woodward (2003) has become one of the most influential approaches to causation in the past decades. It claims that causal relations should be understood in terms of the changes that result from possible interventions: if there is a possible

[7]When binary data is used, network estimation makes use of Ising models, whose edges do not correspond to partial correlations coefficients but can be similarly interpreted.

[8]Statistical tools have been developed that could account for feedback loops in causal graphs, i.e., estimate *directed cyclic graphs* (Spirtes, 1995; Richardson, 1996). However, these techniques have not (yet) been applied to symptom networks, and since their assumptions are stricter than those of DAGs, it is unlikely that these will be met in the context of psychopathology.

intervention on *X* that leads to a change in *Y*, while holding fixed all other variables that could change *Y*, then *X* causes *Y*. A good intervention meets the following criteria:

1. It causes *X*;
2. It acts as a switch for other variables that cause *X*;
3. It does not cause *Y* *via* any other path than *via X*; and
4. It is independent of any variable *Z* which causes *Y* and is on a directed path that does not go through *X* (Woodward, 2003, p. 98).

In this way, interventionism could be used to establish causal relationships between variables without referring to an underlying (neurobiological) common cause: if we demonstrate that an intervention on *X* changes *Y* and does not affect other variables that may cause *X* or *Y*, there is a direct effect of *X* on *Y*. Interventionism thereby allows us to make claims on the relations between variables that go beyond mere correlation. It may not always be empirically possible to construe interventions on symptoms or environmental factors, but this is not necessarily problematic: interventionism requires *hypothetical* interventions that meet the conditions mentioned above (Woodward, 2014, p. 216). So, if (hypothetical) interventions on symptoms or environmental factors can be construed which adhere to Woodward's criteria, we can make causal claims. But are we actually able to come up with (hypothetical) interventions on psychiatric symptoms or environmental factors that adhere to these criteria? In other words, can the network theory meet all assumptions necessary to draw causal conclusions?

If we focus on symptom networks, we see that this may not be as easy as posed. First, it is uncertain whether we can truly eliminate the possibility of a common cause in symptom networks, for this requires us to know (and include) all factors that are casually related to the disorder. If not, it is possible that the causal relation is ultimately due to confounding. If we knew all relevant causal variables, we would still be left with a second problem: it is uncertain whether we can come up with *surgical* hypothetical interventions on symptoms, i.e., interventions that do not influence other variables in the network. Are we able to intervene on a symptom, while keeping other variables in the network stable? It is likely that many symptom interventions have effects on *Y* which do not go through *X* (violation of criterion 3) or influence a variable *Z*, which causes *Y* and is not on a directed path through *X* (violation of criterion 4; Romero, 2015). For example, a peer support group may not be a good surgical intervention to assess whether using medication causes a stable mood, because the peer support group may enhance one's motivation to use medication, but may also facilitate participation in meaningful activities and interaction with helpful group members, which could influence one's mood.[9] One could solve this problem by allowing for *fat-handed* rather than surgical interventions, i.e., interventions that not only affect *X* and other variables on the route from *X* to *Y* but also affect variables affecting *Y* which are not on this route (Woodward, 2008, p. 209; Eberhardt, 2014; Romero, 2015). But even if we allow for this, a third question arises: can we actually take for granted that psychiatric symptoms are distinct and non-overlapping entities? It is necessary to properly define target variables in order to perform suitable interventions. Although proponents of the network theory assume that symptoms are defined at the right level of detail and specificity[10] and "successfully identify the important components in the psychopathology network" (Borsboom, 2017, p. 7), it has also been argued that it is difficult to actually pinpoint individual mental states as suitable targets for intervention (Woodward, 2014). For example, there may be conceptual overlap between the MDD symptoms "depressed mood and diminished pleasure." This is problematic for the application of interventionism to symptom networks: if we are unable to clearly differentiate between two symptoms, we cannot come up with an intervention that does not directly affect both.[11] Lastly, although interventionism could account for networks that are *acyclic*, it is likely that in real life, symptoms influence each other via *feedback loops*. For example, a feedback loop may be present between insomnia, fatigue, concentration problems, and stress (insomnia causes fatigue, which causes concentration problems, which causes stress, which causes insomnia, etc.). If this would be the case, an intervention on the relation between insomnia and fatigue does not act as a switch for concentration problems and stress, thereby violating criterion 2. It may sometimes be possible to circumvent this problem by taking the temporal relations between factors into account (Dijkstra and de Bruin, 2016), but these relations are not always easy to discern. Relatedly, it is possible that symptoms are just too dependent on each other to discern their individual contributions, which hampers our ability to make claims on their individual causal contributions (this will be discussed in more detail in the next section). Hence, although proponents of the network theory argue for an interventionist interpretation of causality, the interventionist criteria which should be satisfied to call a relationship between symptoms causal cannot always be met and/or tested.

What happens when we evaluate the proposed causal relations between environmental factors and symptoms in the network theory in light of the interventionist criteria? First, as discussed previously, proponents of the network theory claim that environmental factors could unidirectionally cause symptoms and thereby serve as catalysts or background elements of the symptom network. It may be possible that such a unidirectional effect can be established more easily for some environmental factors than for individual symptoms. Indeed, for some environmental factors, it may be possible to establish the temporal order of events. For example, when we want to include adverse life events in a psychiatric disorder network, we know

[9]This example was taken from de Bruin (2020).

[10]Borsboom (2017) uses the term "granularity" rather than detail and specificity, but we assume that this was implied.

[11]Interestingly, Woodward (2014) argues that multiple realizability of psychiatric symptoms (i.e., the notion that they may be realized by multiple different physical and/or neural states) could be problematic for applying interventionism to psychiatric disorders, whereas Borsboom et al. (2019a) use multiple realizability as an argument against the traditional view of psychiatric disorders (since it would hamper the possibility of reducing symptoms to brain states).

in some instances that they happened *before* the present-day symptoms arose. This example may run into similar problems of meeting the criteria for good interventions, however. Can we ascertain that we know all relevant causal factors, and can we ensure that (hypothetically) intervening on an environmental factor affects one symptom only? Again, removing people from a stressful home environment may, for example, affect their mood and their agitation. We could circumvent this problem if we allow for fat-handed interventions that influence more than one variable. Can we also do this for the second causal claim made by proponents of the network theory, i.e., that environmental factors could have a direct causal impact on symptom-symptom relations? This claim is more difficult to defend, since intervening on a symptom-symptom relation would likely lead to changes in both symptoms. So, for environmental factors that are clearly temporally distinguishable from the onset of symptoms and under some interpretations of interventionism, we could potentially establish a causal relation between environmental factors and symptoms.

In response, proponents of the network theory could still explain psychiatric disorders as a system of interacting symptoms by referring to the sense-making nature of causal relations. What this section demonstrates, however, is that certain criteria should be met when trying to argue for causal relations in the network theory using interventionism. Whereas these criteria may be met for some effects of environmental factors on symptoms (given certain assumptions), it may be more difficult for others and for symptom-symptom relations. This may limit the potential of the theory to guide psychiatric practice: if it cannot provide evidence for the causal relations underlying psychiatric disorders, it limits their potential to guide psychiatric interventions. But as mentioned previously, Borsboom and colleagues also refer to constitution relations and mechanisms when describing how symptoms and environmental factors relate to psychiatric disorders. Can the network theory provide mechanistic explanations?

Mechanistic Explanation

Mechanistic explanations are concerned with the representation of the mechanisms underlying a certain phenomenon or system, i.e., the phenomenon's components, the components' operations, and their causal organization (Craver and Kaplan, 2018). A mechanistic explanation of chemical neurotransmission, for example, appeals to entities (or components such as ions, neurotransmitters, vesicles, and membranes) and operations (or activities such as depolarizing, diffusing, priming, docking, and fusing) organized together so that they do something – in this case, reliably preserve a signal across the space between cells (Piccinini and Craver, 2011). Mechanistic explanation is the main explanatory strategy in the life sciences, but it does not necessarily go hand in hand with the traditional, reductionist view of psychiatric disorders. Although one could point out that mechanistic explanation is reductionist insofar as it appeals to entities and operations at a lower level of organization, mechanistic explanation does not advocate a sole focus on neurobiology. Indeed, mechanistic explanation typically involves multiple levels of organization and it does not privilege the lowest level. This means that the network theory is theoretically compatible with the mechanistic explanatory strategy, even if it does not include (neuro)biological information.[12]

Can we conceptualize environmental factors as constitutive parts of the mechanism underlying psychiatric disorders? To address this question, we can refer to discussions on the possible extension of cognitive phenomena. Some philosophers have argued that cognitive mechanisms are situated in and dependent on the environment, but that we should not consider environmental factors as part of the mechanism explaining cognitive phenomena. For example, Bechtel (2009, p. 156) states that "for mental phenomena it is appropriate to treat the mind/brain as the locus of the responsible mechanism and to emphasize the boundary between the mind/brain and the rest of the body and between the cognitive agent and its environment." However, Craver (2007, p. 141) suggests that "many cognitive mechanisms draw upon resources outside of the brain and outside of the body to such an extent that it is not fruitful to see the skin, or surface of the central nervous system (CNS), as a useful boundary." If we extrapolate this to psychiatric disorders, we could argue that defining them in an extended sense so that they include brain, body, and environment, allows us to explain them using extended mechanisms.

But if we argue that environmental factors and symptoms can together constitute psychiatric disorders, a different problem arises: where to draw the boundary of the disorder and the mechanism that we want to describe? Recall the example by Borsboom et al. (2019a), in which they state that gambling machines are literally part of the mechanism that explains gambling disorder. If gambling machines are part of this mechanism, why should the mechanism not also include other external entities or events, such as gambling legislation, entry tickets, or socio-cultural norms regarding gambling? Similar claims can be made for substance use disorders. Having an opioid use disorder, for example, depends heavily on the availability of opioids, but does this mean that the person who provides these drugs should be considered part of the disorder's mechanism? These examples illustrate that claiming that environmental factors are a part of the mechanism of a psychiatric disorder raises questions on the *boundaries* of the disorder: where do we draw the line between factors that are explicitly part of the mechanism and thus constitutive for the phenomenon that we want to explain and other external factors that simply causally influence the mechanism or are preconditions for the mechanism's emergence? Craver (2007) has proposed *mutual manipulability* as a criterion to decide whether a part or its activity is constitutively relevant for a phenomenon. According to this criterion, the behavior of a spatiotemporal part X of a system S is constitutively relevant to S's behavior if, and only if, the behaviors of X and S can be mutually manipulated. Craver defines manipulability in

[12]Some may argue that network theory is not compatible with mechanistic explanation because of its "flatness": mechanistic explanations require the presence of multiple layers, but the network theory does not explicate this. We will further address this notion in section "A multilayer network account of psychiatric disorders."

terms of a change in behavior brought about by an intervention à la Woodward (2003). This demarcation criterion is attractive because it could potentially transform the philosophical debate about cognitive extension into a tractable, empirical debate (Kaplan, 2012). However, several philosophers have argued that Craver's mutual manipulability condition is problematic insofar as it undermines the fundamental distinction between constitution and causation. Indeed, constitution is typically treated as a *non-causal* dependency relation between lower-level parts and higher-level mechanisms. This issue is still a subject of intense debate. To provide a definition of constitutive relationships in terms of interventionism, some have argued for the use of the fat-handed intervention criterion (Romero, 2015; Baumgartner and Gebharter, 2016; Baumgartner and Casini, 2017). Nevertheless, as demonstrated earlier, interpreting network theory in light of (fat-handed) interventionism still faces important challenges, hampering the possibility to establish mutual manipulability relations using interventionism. Hence, it is uncertain whether adding this demarcation criterion would help to decide the issue in the context of the network theory.

There is another, more pressing problem for the mechanistic explanatory potential of the network theory: in order to construe a mechanistic explanation of a phenomenon, the phenomenon should be *decomposable* in terms of components (structural decomposition) and operations (functional decomposition). Recall the example on chemical neurotransmission: this phenomenon is mechanistically explanatory because it is structurally decomposable in terms of ions, neurotransmitters, vesicles, and membranes, and functionally decomposable in terms of depolarization, diffusion, priming, docking, and fusion. Are psychiatric disorders decomposable in this sense? It has been argued that there are two types of systems with different levels of decomposability. In a *nearly decomposable* system, the behavior of the system's individual components is integrated, but the components can still be understood and studied independently. Bechtel (2009) argues that cognitive systems are nearly decomposable, meaning that they can be explained mechanistically. In a *non-decomposable* system, the (short-term) behavior of the system's component parts highly depends on the behavior of other individual component parts. Since no subsystems of components are (nearly) independent of one another, the system cannot be explained mechanistically (Rathkopf, 2018). It is an open-ended question which system best describes psychiatric disorders. It may be possible that psychiatric disorders are in fact nearly decomposable, and that the theory's current description of psychiatric disorders in terms of symptoms and environmental factors provides a mechanism sketch that can be filled in with more (structural) details as more research becomes available (Piccinini and Craver, 2011). However, it may also be possible that psychiatric disorders are in fact non-decomposable. As mentioned earlier, the network theory claims that symptoms operate in causal feedback loops. If systems are characterized by circular causality, i.e., a given component of the system is both continuously affecting and simultaneously being affected by activity in another component, it is difficult to identify the contribution of the component in question in terms of the underlying structural entities (Lamb and Chemero, 2014).[13] Even if this were possible, we still face the problem discussed previously: individual symptoms may not be as easily differentiated as commonly assumed, thereby limiting the decomposability of psychiatric disorders. If we conclude on the basis of these considerations that psychiatric disorders are in fact non-decomposable systems, we cannot explain them mechanistically and cannot substantiate the claims made by proponents of the network theory concerning constitution.

This section addressed two issues concerning the mechanistic explanatory potential of the network theory. First, we showed that there are difficulties in justifying that environmental factors *constitute* or *cause* psychiatric disorders or symptoms. Second, we can only substantiate the claim that symptoms and environmental factors co-constitute psychiatric disorders using mechanistic explanation if psychiatric disorders are in fact decomposable.[14] This does not imply that the network theory cannot help us to explain the development and guide the treatment of psychiatric disorders. Rather, it demonstrates that it can only have mechanistic explanatory potential when certain criteria are met, and when we adopt a specific understanding of mechanistic explanation.

THE TOPOLOGICAL EXPLANATORY STRATEGY

Proponents of the network theory do not only make reference to the individual relations between factors, but also to the characteristics of symptom networks themselves. Borsboom (2017, p. 7) argues, for instance, that the psychopathology network, an interdiagnostic network including all possible psychiatric symptoms, "has a non-trivial topology, in which certain symptoms are more tightly connected than others. These symptom groupings give rise to the phenomenological manifestation of mental disorders as groups of symptoms that often arise together." The psychopathology network thus features *clustering*, i.e., groups of strongly related nodes (Borsboom et al., 2011). However, it is also suggested that the characteristics of symptom networks can explain the development and sustenance of psychiatric disorders. Indeed, Borsboom (2017) argues that the presence of high symptom-symptom connectivity can explain the dynamics of psychiatric disorders: in symptom networks with *high connectivity*, symptoms continue to activate each other after the initial activation of one symptom. Is this claim compatible with a topological explanatory strategy?

[13]Note that the concept of circular causality itself has received criticism (Bakker, 2005).

[14]One could argue that (structural) decomposition is not essential for mechanistic explanation (Zednik, 2014), and that it is more important that mechanistic explanations demonstrate how phenomena are "situated in the causal structure of the world" (Craver, 2013, p. 134). However, as argued previously, demonstrating causal relationships in the context of network theory may also pose issues.

Topological explanations explain the dynamics of complex systems by making use of topological properties, i.e., properties of a complex system that are mathematically quantified using graph theory (Kostić, 2019). To illustrate what topological properties are, a classic example might help. In their seminal publication, Watts and Strogatz (1998) used networks to examine, among others, how infectious diseases spread by studying two topological properties: the *characteristic path length* and the *clustering coefficient*. Path length refers to the number of edges (i.e., the graph-theoretical term for relations) on the shortest path between two nodes (i.e., the graph-theoretical term for variables), and the characteristic path length is defined as the average shortest path length between all pairs of nodes in the network. The clustering coefficient is a measure of the cliquishness of the network (i.e., the degree to which nodes near each other are strongly connected). Watts and Strogatz (1998) discovered empirically that many networks have high clustering coefficients and short characteristic path lengths, a topological property they called the *small-world property*. Their simulations demonstrated that the human population is like a small-world network, which explains why diseases can spread quickly throughout the population.

This example illustrates that topological properties can be used to explain the dynamics of a system constituted by interacting parts. But what exactly is meant by *explain* in this context?[15] According to Kostić (2020), a topological explanation supports counterfactuals that describe a counterfactual dependency between a system's topological properties and its network dynamics (i.e., if the topological property would not have been there, the network dynamics would have been different). He distinguishes two ways in which topological explanations may describe counterfactual dependency relations: a *vertical* explanation in which a global topological property (characteristic of the whole network) determines certain general properties of the real-world system, and a *horizontal* explanation in which a local topological property (characteristic of a part of the network) determines certain local dynamical properties of the real-world system. Kostić (2020) illustrates the difference between these two modes of explanation by focusing on the question of cognitive control, i.e., how the brain as a dynamical system efficiently transitions between internal states. If the explanation-seeking question is: "why can the brain achieve cognitive control?," the relevant vertical counterfactual is: if the brain would not have been a small-world network, it would not have been able to achieve cognitive control. If the explanation-seeking question is: "how and why can the brain efficiently transition between states?," one of the relevant horizontal counterfactuals is: had the local topological properties not determined the energy requirements for those transitions, then these energy requirements would have been different. How can counterfactual dependence account for explanatory asymmetry, i.e., the topological property explaining the phenomenon and not vice versa? Kostić (2020) suggests three ways in which this can be done. First, the phenomenon that the topological property wants to explain is not a mathematically quantified property, hence there is *property asymmetry*. Second, there is *counterfactual asymmetry*: the phenomenon depends on the topological property, but the topological property does not depend on the phenomenon. Third, reversing the direction of explanation makes the claim non-explanatory. If the explanation-seeking question is: "why does a system have a certain topological property," referring to the phenomenon is not a scientifically relevant answer. Hence, there is *perspectival asymmetry*.

The claim by Borsboom (2017) concerning connectivity can be interpreted as a vertical topological explanation: a global, mathematically quantifiable property of the network (i.e., high connectivity) explains the vulnerability to develop a psychiatric disorder. If symptoms would be less strongly connected, one would be less vulnerable to developing a psychiatric disorder. Support for this counterfactual dependency has been provided by network analysis. Indeed, Borsboom (2017) refers to a within-subject study demonstrating that in MDD, altering a parameter that determines symptom network connectivity changes the network's vulnerability: when the nodes are highly connected, this increases the likelihood that activation of one symptom leads to activation of other symptoms, making it less likely for these symptoms to disappear (Cramer et al., 2016). Relatedly, high symptom network connectivity in MDD has also been associated with having a persistent diagnosis after 2 years (van Borkulo et al., 2015). So, it is possible for the network theory to make use of topological properties that counterfactually explain the dynamics of a psychiatric disorder.

An appealing feature of topological explanations is that they can and should be used to provide explanations of non-decomposable systems (Rathkopf, 2018). To illustrate this, Rathkopf uses the topological property *edge betweenness*, i.e., the number of the shortest paths between pairs of nodes that go through that specific edge (Girvan and Newman, 2002). Betweenness is a measure of the extent to which an edge occupies a central place in the network. To compute the betweenness of an edge, the shortest path length between all pairs of nodes in the network is examined, after which it is calculated what proportion of those paths incorporate that edge. This means that betweenness applies to a single edge, but that its value indirectly refers to the rest of the graph. In this way, it combines the complex patterns of interaction into one meaningful variable with explanatory power, making the non-decomposable system "epistemically accessible" (Rathkopf, 2018, p. 72). In other words, topological explanations can provide meaningful insights into psychiatric disorders if we are not able to clearly differentiate (the activity of) their underlying components.

The topological explanatory strategy does pose some difficulties in the context of psychiatric disorders, however. First, providing the right topological explanations depends on the topological property (and the phenomenon it aims to explain) to be "approximately true" (Kostić, 2020, p. 2). We can estimate topological properties using network analysis, but as highlighted

[15]Some philosophers have questioned the explanatory potential of topological properties. For example, Craver (2016) argues that topological explanations are in fact exploratory, because they cannot distinguish good from bad explanations. Moreover, as an anonymous reviewer pointed out, one could argue that topological explanations do not provide information on *why* certain topological properties, and not a relevant contrast class, yield these network dynamics.

previously, we should critically examine the data and statistical methods used to substantiate theoretical claims. Second, it is not always clear what the relevant counterfactuals are for a topological explanation: would a relevant counterfactual be an instance in which the psychiatric disorder is not present at all, or if symptom severity is decreased, for example? Third, how to interpret the global and local topological properties we discover is not always straightforward. For example, a set of topological properties that is frequently examined in the context of symptom networks is measures related to *centrality*. These measures reveal the relative importance of nodes in a network structure. It has been argued, however, that they may not have meaningful interpretations in the context of psychiatry, because they come with assumptions that are not necessarily met in psychopathological networks (Bringmann et al., 2019). This especially concerns global centrality measures that depend on the network as a whole (e.g., betweenness and closeness centrality).

A final issue is that thus far, we have only focused on topological explanations of the symptom network. How could the network theory include environmental factors in its topological explanations of psychiatric disorders? One option is to assess the dynamics of the symptom network with and without the presence of a certain environmental factor (e.g., Choi et al., 2017; Hasmi et al., 2018). This option, however, only allows one to make claims on the role of an environmental factor on the symptom network as a whole, and does not suffice when we are interested in multiple environmental factors (that we do not want to average) and their interactions. Alternatively, we could include environmental factors as part of a network structure. The next section will present the *multilayer* network account of psychiatric disorders as a framework for the network theory that could accommodate topological *and* causal/mechanistic explanatory strategies.

A MULTILAYER NETWORK ACCOUNT OF PSYCHIATRIC DISORDERS

The network theory may benefit from explicitly adopting a multilayer network account of psychiatric disorders. A multilayer network can be defined as a network of networks, or a network that is comprised of multiple layers with connections between and within the layers. In recent years, statistical techniques have been developed that allow for the estimation of such networks (Kivelä et al., 2014). Multilayer networks have been used to study various complex phenomena, including social, biological, and transport systems (Mucha et al., 2010; Boccaletti et al., 2014; De Domenico et al., 2014, 2016). They are also increasingly used in network neuroscience to integrate different neuroimaging modalities (e.g., to compare the structural and functional connectivity of brain regions), or to study brain networks over different time points, among others (De Domenico, 2017; Vaiana and Muldoon, 2018). What provides these networks with an advantage over *monolayer* networks is that the latter often require data to be aggregated (for example, by means of averaging) or to be ignored. Multilayer networks can retain this information by including it in different layers, making them better suited to deal with multidimensional data and allowing for analyses that could not be performed when focusing on one layer of analysis only.

Researchers have suggested that multilayer networks should also be applied to the study of psychiatric disorders (Braun et al., 2018). However, multilayer network analysis typically requires nodes to be replicated over the different layers, which poses a problem if we want to integrate information from different scales (e.g., symptoms and environmental factors) as layers in the multilayer network structure. Fortunately, statistical techniques are available that do not require such node replication (Brooks et al., 2020). This enables the statistical estimation of multilayer networks including various different factors that are relevant to the development, sustenance and potential treatment of psychiatric disorders. It has been argued that these innovations in multilayer network analysis techniques should be paired with innovations in the *theoretical* frameworks of psychiatric disorders, doing justice to their dimensional and multiplex nature (Braun et al., 2018).

Although proponents of the network theory do not explicitly endorse a multilayer network account of psychiatric disorders, their claims are compatible with this view. More specifically, the multilayer network account provides an explicit framework for the network theory that can include multiple different factors, with the additional advantage that it can be statistically modeled.[16] First, it is compatible with the claim that "basically every element of the system is dependent on a heterogeneous set of biological and external factors" (Borsboom et al., 2019a, p. 9). Multilayer networks provide a framework that can easily be extended to accommodate various non-symptom factors interacting with the symptom network. Second, proponents of the network theory claim that environmental factors could be part of the *mechanism* that *constitutes* symptoms or symptom-symptom relations. It may be possible for multilayer networks to account for this claim when symptoms and environmental factors are construed as different layers in the network structure.

A multilayer network account has other explanatory advantages as well, insofar as it might be able to accommodate both mechanistic/causal explanations and topological explanations of psychiatric disorders. First, a multilayer network account may enhance the mechanistic explanatory potential of the network theory, by incorporating different factors that are part of the mechanisms underlying psychiatric disorders. In this sense, the account is compatible with the claim that psychiatric disorders are *mechanistic property clusters*: clusters of properties that span multiple layers and are maintained by interacting, dysfunctional,

[16]Interestingly, proponents of the network theory seem sympathetic to the idea that different factors related to psychiatric disorders may represent different network structures. Borsboom et al. (2019b) argue that psychological networks may relate to underlying biological networks, either in a part-whole relationship or with biological networks being *nested* in a symptom network. This latter statement is similar to a claim made in an earlier article, stating that "the reality of psychopathology involves a Russian doll of networks nested within networks in several layers of complexity" (Borsboom and Cramer, 2013, p. 104). Here, they argue that symptom networks could relate to networks of environmental factors (i.e., social networks) and to neurobiological networks. However, their suggestions present methodological difficulties, as it is not clear how nested networks could be modeled statistically (Borsboom et al., 2019b).

and self-sustaining mechanisms (Kendler et al., 2011). Both accounts argue that there is not one layer that can tell us all we want to know about a psychiatric disorder: rather, complex and multi-layer causal mechanisms, including genetic, cellular, neural, psychological, environmental and socio-cultural factors produce, underlie and sustain psychiatric disorders (Kendler, 2008). However, as claimed earlier, psychiatric disorders can only be explained mechanistically if they are decomposable. Multilayer networks could include layers with a higher degree of decomposability, such as structural neurobiology (e.g., anatomical connectivity obtained with diffusion-weighted magnetic resonance imaging). Since such an underlying network could include information on concrete parts and operations (and their causal interactions), it would allow for the possibility of structural decomposition as is required by the mechanistic explanatory strategy. Structural data could also constrain functional data (e.g., Suárez et al., 2020), compatible with the mechanistic claim that function needs to be constrained by structure. One could also speculate that these layers with higher decomposability may meet more of the criteria for good interventions than purely functional layers, which means that their inclusion could allow for local causal explanations of elements of psychiatric disorders. So, a multilayer network account may enhance the mechanistic explanatory potential of the network theory, although this hinges on the issue of the decomposability of psychiatric disorders and the layers that such a theoretical framework would incorporate.

However, the multilayer network account could also enhance the explanatory potential of the network theory if psychiatric disorders turn out to be non-decomposable. More specifically, it allows for topological explanations that go beyond symptom networks. In this way, it can do justice to the idea that interactions between non-symptom factors are relevant for explaining the development, sustenance, or potential treatment of psychiatric disorders. First, topological properties of non-symptom layers may inform us about the topological properties of the symptom network. As mentioned above, high connectivity between symptoms has been related to increased vulnerability to develop psychiatric disorders. Psychiatric disorder-related changes in connectivity patterns have also been demonstrated in networks at multiple layers of brain organization (van den Heuvel and Sporns, 2019; van den Heuvel et al., 2019). So, exploring the dynamics of non-symptom layers of the multilayer network structure may provide information about the dynamics of the symptom network. Second, multilayer networks may allow for topological explanations of psychiatric disorders that span multiple layers. Indeed, statistical techniques have both extended traditional topological properties to multilayer networks and developed topological properties specific to multilayer structures (see Vaiana and Muldoon, 2018 for an overview). Such multilayer topological explanations may provide new insights into the dynamics of psychiatric disorders that supersede what we could explain if we solely focus on the symptom network. For example, De Domenico et al. (2015) demonstrated that hubs in multilayer neural networks differ dramatically from hubs in separate layers of the system, and Battiston et al. (2014) showed that two layers in a multilayer network exhibited different network properties but shared certain hubs and motifs (i.e., characteristic recurrent connection patterns). What could a multilayer topological explanation look like in the context of psychiatric disorders? A topological property that could be exploited is *community structure*, i.e., the presence of groups of nodes with strong internal and weak external connections. If time is added as a dimension to the multilayer network structure, the dynamical changes in community structure over time could be investigated. Braun et al. (2018) have suggested that this could be applied to the study of brain networks in individuals with a psychiatric disorder diagnosis to identify possible critical time points in their clinical development. In a similar fashion, examining how the (community structure of) the symptom network changes over time may explain the development of psychiatric disorders. Moreover, multilayer topological properties could be used to investigate and explain heterogeneity within psychiatric disorders by identifying subtypes with different multilayer topologies (e.g., including different symptoms and neurobiological factors; similar to a suggestion in the context of personality research Brooks et al., 2020).

This section demonstrated that adopting a multilayer network account could allow the network theory to accommodate both mechanistic/causal and topological explanations of psychiatric disorders spanning multiple layers. On such an account, the explanatory potential of the network theory does not hinge on whether psychiatric disorders are (nearly) decomposable. If psychiatric disorders or a specific layer turn out to be non-decomposable, it may still be possible to account for their dynamics using topological explanations, meaning that a multilayer network account is able to address a variety of explanation-seeking questions. Of course, more statistical and conceptual research into multilayer networks of psychiatric disorders is necessary to further explore their potential. Future research could, for example, examine which layers and relations are relevant to include in consultation with clinicians and experts by experience. Also, it should be examined how different layers can be defined, how they relate to each other, and which statistical methods would be most suited to estimate such networks using empirical datasets. Lastly, it should be assessed how a multilayer network account can translate to clinical practice, and to what extend it is compatible with existing theoretical frameworks (such as RDoC, with different domains potentially being represented as different layers of a multilayer network).

CONCLUSION

This article critically examined the explanatory potential of the network theory that includes both symptoms and environmental factors. On the one hand, proponents of the network theory claim that causally interacting symptoms constitute psychiatric disorders and that environmental factors causally and mechanistically influence symptoms and psychiatric disorders in general. This suggests that the network theory could provide causal/mechanistic explanations of psychiatric disorders. However, to justify these

claims, various assumptions should be satisfied. We cannot make causal claims based on network analysis alone, and determining causality using Woodward's interventionist theory requires psychiatric disorders and their symptoms to meet criteria for suitable interventions, which may not always be possible. Moreover, providing a mechanistic account of psychiatric disorders is only possible if they are decomposable, and even then may it be difficult to formally differentiate between causal and constitutive relations. On the other hand, proponents of the network theory suggest that it might be possible to explain psychiatric disorders in terms of the characteristics of symptom networks themselves. We showed that adopting a topological explanatory strategy may be promising for the network theory, for it can explain the dynamics of psychiatric disorders when they are non-decomposable, but it does pose issues as well. Lastly, we argue that adopting a multilayer network account of psychiatric disorders provides a framework for the network theory that could accommodate different factors related to psychiatric disorders as well as both mechanistic/causal and topological explanations.

A multilayer network account differs vastly from the traditional view of psychiatric disorders we started with. Critical voices may argue that we have traded a relatively straightforward account of how to understand and explain psychiatric disorders with an overly complex alternative. Indeed, arguing that psychiatric disorders are brain disorders seems much easier than appealing to an account of psychiatric disorders that includes different types of factors and relations between layers and individual factors. However, it is unlikely that our explanations of psychiatric disorders will ultimately be simple (as demonstrated by the lack of empirical support for the traditional view). Instead of trying to reduce the complexity of psychiatric disorders, it may be preferable to embrace their complex and multifaceted nature. An account that does this while still having explanatory potential may ultimately provide a more comprehensive understanding of psychiatric disorders and more guidance for psychiatric practice. The network theory should be applauded for aiming to provide an explanatory framework that captures some of this complexity, and the multilayer network account should be seen as a possible elaboration of this theory. This does not mean that the multilayer network account is the only conceptualization of psychiatric disorder that does justice to their complexity. Nonetheless, moving toward such an account may be more fruitful for psychiatry than moving toward oversimplification.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

NB is responsible for the final structure of the manuscript and has primarily contributed to the sections on the interpretation of the network theory, network analysis, Woodward's interventionist theory, topological explanations, and multilayer networks. LB has primarily contributed to the sections on mechanistic and topological explanations (in multilayer networks). GG and JG have provided supervision and feedback on the argumentation. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank the peer reviewers for their comments and Freek Oude Maatman and Marc Slors for the helpful feedback.

REFERENCES

Adam, D. (2013). On the spectrum. *Nature* 496, 416–418. doi: 10.1038/496416a

American Psychological Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-V). 5th Edn*. Washington, DC: American Psychiatric Publishing.

Bakker, B. (2005). The concept of circular causality should be discarded. *Behav. Brain Sci.* 28, 195–196. doi: 10.1017/S0140525X05230042

Barabási, A. -L. (2012). The network takeover. *Nat. Phys.* 8, 14–16. doi: 10.1038/nphys2188

Battiston, F., Nicosia, V., and Latora, V. (2014). Structural measures for multiplex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 89, 1–14. doi: 10.1103/PhysRevE.89.032804

Baumgartner, M., and Casini, L. (2017). An abductive theory of constitution. *Philos. Sci.* 84, 214–233. doi: 10.1086/690716

Baumgartner, M., and Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *Br. J. Philos. Sci.* 67, 731–756. doi: 10.1093/bjps/axv003

Beard, C., Millner, A. J., Forgeard, M. J. C., Fried, E. I., Hsu, K. J., Treadway, M., et al. (2016). Network analysis of depression and anxiety symptom relations in a psychiatric sample. *Psychol. Med.* 46, 3359–3369. doi: 10.1016/j.physbeh.2017.03.040

Bechtel, W. (2009). "Explanation: mechanism, modularity, and situated cognition" in *Cambridge handbook of situated cognition*. eds. P. Robbins and M. Aydede (Cambridge: Cambridge University Press), 155–170.

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., et al. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* 544, 1–122. doi: 10.1016/j.physrep.2014.07.001

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *J. Clin. Psychol.* 64, 1089–1108. doi: 10.1002/jclp.20503

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry* 16, 5–13. doi: 10.1002/wps.20375

Borsboom, D., and Cramer, A. O. J. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annu. Rev. Clin. Psychol.* 9, 91–121. doi: 10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Cramer, A. O. J., and Kalis, A. (2019a). Brain disorders? Not really: why network structures block reductionism in psychopathology research. *Behav. Brain Sci.* 42, 1–63. doi: 10.1017/S0140525X17002266

Borsboom, D., Cramer, A. O. J., and Kalis, A. (2019b). Reductionism in retreat. *Behav. Brain Sci.* 42:e32. doi: 10.1017/S0140525X18002091

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., and Waldorp, L. J. (2011). The small world of psychopathology. *PLoS One* 6:e27407. doi: 10.1371/journal.pone.0027407

Braun, U., Schaefer, A., Betzel, R. F., Tost, H., Meyer-Lindenberg, A., and Bassett, D. S. (2018). From maps to multi-dimensional network mechanisms of mental disorders. *Neuron* 97, 14–31. doi: 10.1016/j.neuron.2017.11.007

Bringmann, L. F., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., and Wigman, J. T. W. (2019). What do centrality measures measure in psychological networks? *J. Abnorm. Psychol.* 128, 892–903. doi: 10.1037/abn0000446

Bringmann, L. F., and Eronen, M. I. (2018). Don't blame the model: reconsidering the network approach to psychopathology. *Psychol. Rev.* 125, 606–615. doi: 10.1037/rev0000108

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., et al. (2013). A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS One* 8:e60188. doi: 10.1371/journal.pone.0060188

Brooks, D., Hulst, H. E., de Bruin, L., Glas, G., Geurts, J. J. G., and Douw, L. (2020). The multilayer network approach in the study of personality neuroscience. *Brain Sci.* 10:915. doi: 10.3390/brainsci10120915

Choi, K. W., Batchelder, A. W., Ehlinger, P. P., Safren, S. A., and O'Cleirigh, C. (2017). Applying network analysis to psychological comorbidity and health behavior: depression, PTSD, and sexual risk in sexual minority men with trauma histories. *J. Consult. Clin. Psychol.* 85, 1158–1170. doi: 10.1037/ccp0000241

Colombo, M., and Heinz, A. (2019). Explanatory integration, computational phenotypes, and dimensional psychiatry: the case of alcohol use disorder. *Theor. Psychol.* 29, 697–718. doi: 10.1177/0959354319867392

Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., et al. (2016). Major depression as a complex dynamic system. *PLoS One* 11:e0167490. doi: 10.1371/journal.pone.0167490

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press.

Craver, C. F. (2013). "Functions and mechanisms: a perspectivalist view" in *Functions: Selection and mechanisms*. ed. P. Huneman (Dordrecht: Springer), 133–158.

Craver, C. F. (2016). The explanatory power of network models. *Philos. Sci.* 83, 698–709. doi: 10.1086/687856

Craver, C. F., and Kaplan, D. M. (2018). Are more details better? On the norms of completeness for mechanistic explanations. *Br. J. Philos. Sci.* 71, 1–33. doi: 10.1093/bjps/axy015

de Bruin, L. (2020). Managing the self: some philosophical issues. *Philos. Psychiatry Psychol.* 27, 371–373. doi: 10.1353/ppp.2020.0047

De Domenico, M. (2017). Multilayer modeling and analysis of human brain networks. *Gigascience* 6, 1–8. doi: 10.1093/gigascience/gix004

De Domenico, M., Granell, C., Porter, M. A., and Arenas, A. (2016). The physics of spreading processes in multilayer networks. *Nat. Phys.* 12, 901–906. doi: 10.1038/nphys3865

De Domenico, M., Solé-Ribalta, A., Gómez, S., and Arenas, A. (2014). Navigability of interconnected networks under random failures. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8351–8356. doi: 10.1073/pnas.1318469111

De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S., and Arenas, A. (2015). Ranking in interconnected multilayer networks reveals versatile nodes. *Nat. Commun.* 6:6868. doi: 10.1038/ncomms7868

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Dijkstra, N., and de Bruin, L. (2016). Cognitive neuroscience and causal inference: implications for psychiatry. *Front. Psychol.* 7:129. doi: 10.3389/fpsyt.2016.00129

Eberhardt, F. (2014). Direct causes and the trouble with soft interventions. *Erkenntnis* 79, 755–777. doi: 10.1007/s10670-013-9552-2

Epskamp, S., Kruis, J., and Marsman, M. (2017). Estimating psychopathological networks: be careful what you wish for. *PLoS One* 12:e0179891. doi: 10.1371/journal.pone.0179891

Fisher, A. J., Medaglia, J. D., and Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proc. Natl. Acad. Sci. U. S. A.* 115, E6106–E6115. doi: 10.1073/pnas.1711978115

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. PsyArxiv [Preprint]. doi: 10.31234/osf.io/zg84s

Fried, E. I., Bockting, C., Arjadi, R., Borsboom, D., Amshoff, M., Cramer, A. O. J., et al. (2015). From loss to loneliness: the relationship between bereavement and depressive symptoms. *J. Abnorm. Psychol.* 124, 256–265. doi: 10.1037/abn0000028

Fried, E. I., and Cramer, A. O. J. (2017). Moving forward: challenges and directions for psychopathological network theory and methodology. *Perspect. Psychol. Sci.* 12, 999–1020. doi: 10.1177/1745691617705892

Fullana, M. A., Abramovitch, A., Via, E., López-Sola, C., Goldberg, X., Reina, N., et al. (2020). Diagnostic biomarkers for obsessive-compulsive disorder: a reasonable quest or ignis fatuus? *Neurosci. Biobehav. Rev.* 118, 504–513. doi: 10.1016/j.neubiorev.2020.08.008

Girvan, M., and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U. S. A.* 99, 7821–7826. doi: 10.1073/pnas.122653799

Hasmi, L., Drukker, M., Guloksuz, S., Viechtbauer, W., Thiery, E., Derom, C., et al. (2018). Genetic and environmental influences on the affective regulation network: a prospective experience sampling analysis. *Front. Psychol.* 9:602. doi: 10.3389/fpsyt.2018.00602

Insel, T., and Cuthbert, B. N. (2015). Brain disorders? Precisely: precision medicine comes to psychiatry. *Science* 348, 499–500. doi: 10.1126/science.aab2358

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* 167, 748–751. doi: 10.1176/appi.ajp.2010.09091379

Isvoranu, A. M., Borsboom, D., Van Os, J., and Guloksuz, S. (2016). A network approach to environmental impact in psychotic disorder: brief theoretical framework. *Schizophr. Bull.* 42, 870–873. doi: 10.1093/schbul/sbw049

Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biol. Philos.* 27, 545–570. doi: 10.1007/s10539-012-9308-4

Kendler, K. S. (2008). Explanatory models for psychiatric illness. *Am. J. Psychiatry* 165, 695–702. doi: 10.1176/appi.ajp.2008.07071061

Kendler, K. S., Zachar, P., and Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychol. Med.* 41, 1143–1150. doi: 10.1017/S0033291710001844

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *J. Complex Netw.* 2, 203–271. doi: 10.1093/comnet/cnu016

Kostić, D. (2019). Minimal structure explanations, scientific understanding and explanatory depth. *Perspect. Sci.* 27, 48–67. doi: 10.1162/posc_a_00299

Kostić, D. (2020). General theory of topological explanations and explanatory asymmetry. *Philos. Trans. R. Soc. B Biol. Sci.* 375:20190321. doi: 10.1098/rstb.2019.0321

Lamb, M., and Chemero, A. (2014). "Structure and application of dynamical models in cognitive science" in *Proceedings of the 36th annual conference of the cognitive science society*; July 26, 2014; 809–814.

Moffa, G., Catone, G., Kuipers, J., Kuipers, E., Freeman, D., Marwaha, S., et al. (2017). Using directed acyclic graphs in epidemiological research in psychosis: an analysis of the role of bullying in psychosis. *Schizophr. Bull.* 43, 1273–1279. doi: 10.1093/schbul/sbx013

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J. -P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328, 876–878. doi: 10.1126/science.1184819

Nolen-Hoeksema, S., and Watkins, E. R. (2011). A heuristic for developing transdiagnostic models of psychopathology: explaining multifinality and divergent trajectories. *Perspect. Psychol. Sci.* 6, 589–609. doi: 10.1177/1745691611419672

Oude Maatman, F. (2020). Reformulating the network theory of mental disorders: folk psychology as a factor, not a fact. *Theor. Psychol.* 30, 703–722. doi: 10.1177/0959354320921464

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, London: Cambridge University Press.

Piccinini, G., and Craver, C. F. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese* 183, 283–311. doi: 10.1007/s11229-011-9898-4

Rathkopf, C. (2018). Network representation and complex systems. *Synthese*, 55–78. doi: 10.1007/s11229-015-0726-0

Richardson, T. S. (1996). "A discovery algorithm for directed cyclis graphs" in *Proceedings of the twelfth international conference on uncertainty in artificial intelligence*; August 1, 1996; 454–461.

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., and Borsboom, D. (2020). The network approach to psychopathology: a review of the literature 2008-2018 and an agenda for future research. *Psychol. Med.* 50, 353–366. doi: 10.1017/S0033291719003404

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese* 192, 3731–3755. doi: 10.1007/s11229-015-0718-0

Slors, M., Francken, J. C., and Strijbos, D. (2019). Intentional content in psychopathologies requires an expanded interpretivism. *Behav. Brain Sci.* 42:e26. doi: 10.1017/S0140525X18001176

Spirtes, P. (1995). "Directed cyclic graphical representations of feedback models" in *Proceedings of the eleventh conference on uncertainty in artificial intelligence;* August 20, 1995; 491–498.

Suárez, L. E., Markello, R. D., Betzel, R. F., and Misic, B. (2020). Linking structure and function in macroscale brain networks. *Trends Cogn. Sci.* 24, 302–315. doi: 10.1016/j.tics.2020.01.008

Vaiana, M., and Muldoon, S. F. (2018). Multilayer brain networks. *J. Nonlinear Sci.* 30, 2147–2169. doi: 10.1007/s00332-017-9436-8

van Borkulo, C., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., Waldorp, L. J., and Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry* 72, 1219–1226. doi: 10.1001/jamapsychiatry.2015.2079

van den Heuvel, M. P., Scholtens, L. H., and Kahn, R. S. (2019). Multiscale neuroscience of psychiatric disorders. *Biol. Psychiatry* 86, 512–522. doi: 10.1016/j.biopsych.2019.05.015

van den Heuvel, M. P., and Sporns, O. (2019). A cross-disorder connectome landscape of brain dysconnectivity. *Nat. Rev. Neurosci.* 20, 435–446. doi: 10.1038/s41583-019-0177-6

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of "small-world" network. *Nature* 393, 440–442. doi: 10.1038/30918

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Woodward, J. (2008). *Invariance, modularity, and all that: Cartwright on causation*. eds. S. Hartman, C. Hoefer and L. Bovens (New York: Routledge).

Woodward, J. (2014). "Cause and explanation in psychiatry: an interventionist perspective" in *Philosophical issues in psychiatry: Explanation, phenomenology and nosology*. eds. K. S. Kendler and J. Parnas (Baltimore: John Hopkins University Press), 209–272.

Zednik, C. (2014). "Are systems neuroscience explanations mechanistic?" in *Preprint volume for philosophy of science association 24th biennial meeting* (Chicago, IL: Philosophy of Science Association), 954–975.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 March 2021
doi: 10.3389/fpsyg.2021.604393

Let Me Make You Happy, and I'll Tell You How You Look Around: Using an Approach-Avoidance Task as an Embodied Emotion Prime in a Free-Viewing Task

Artur Czeszumski[1]*, *Friederike Albers*[1], *Sven Walter*[1] and *Peter König*[1,2]

[1] Institute of Cognitive Science, Universität Osnabrück, Osnabrück, Germany, [2] Institut für Neurophysiologie und Pathophysiologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany

The embodied approach of human cognition suggests that concepts are deeply dependent upon and constrained by an agent's physical body's characteristics, such as performed body movements. In this study, we attempted to broaden previous research on emotional priming, investigating the interaction of emotions and visual exploration. We used the joystick-based approach-avoidance task to influence the emotional states of participants, and subsequently, we presented pictures of news web pages on a computer screen and measured participant's eye movements. As a result, the number of fixations on images increased, the total dwell time increased, and the average saccade length from outside of the images toward the images decreased after the bodily congruent priming phase. The combination of these effects suggests increased attention to web pages' image content after the participants performed bodily congruent actions in the priming phase. Thus, congruent bodily interaction with images in the priming phase fosters visual interaction in the subsequent exploration phase.

Keywords: overt attention, eye-tracking, emotions, cognitive bias modification, automatic approach bias, embodiment, approach avoidance task

OPEN ACCESS

Edited by:
Leon De Bruin, Radboud University Nijmegen, Netherlands

Reviewed by:
Fernando Marmolejo-Ramos, University of South Australia, Australia
Ophelia Deroy, Ludwig Maximilian University of Munich, Germany

***Correspondence:**
Artur Czeszumski
aczeszumski@uos.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *09 September 2020*
Accepted: *18 February 2021*
Published: *15 March 2021*

Citation:
Czeszumski A, Albers F, Walter S and König P (2021) Let Me Make You Happy, and I'll Tell You How You Look Around: Using an Approach-Avoidance Task as an Embodied Emotion Prime in a Free-Viewing Task. Front. Psychol. 12:604393. doi: 10.3389/fpsyg.2021.604393

INTRODUCTION

Gaze-dependent shifts play a pivotal role in visual processing. Using modern eye-tracking techniques, it is possible to measure overt shifts of attention reliably and unobtrusively, helping us understand eye movement behavior. What one observes is influenced by at least three factors. First, attention is influenced by the external stimuli' properties, processed in a bottom-up hierarchy (Treisman and Gelade, 1980; Itti and Koch, 2000). This includes low-level features of the visual stimulus, for instance, contrast, contours, color, texture, and motion. However, it may also include more complex features like complex shapes of objects or the emotional valence of images (Thomas and Hasher, 2006; Einhäuser et al., 2008). Second, attention is influenced by internal variables like task-demands (Hayhoe et al., 2003; Einhäuser and Koch, 2008; Rothkopf et al., 2016), as well as the observer's emotional state (Kaspar et al., 2013). Third, the spatial factors like the central bias (Tatler, 2007) and saccadic momentum (Wilming et al., 2013) influence the selection of fixation targets. These three factors' relative contribution is a matter of debate (Kollmorgen et al., 2010), and presumably depends on the precise circumstances (Einhäuser and Koch, 2008). Additionally, to all

different mentioned levels that attention can be influenced, it is crucial to operationalize attention itself (Hommel et al., 2019). Our study used direction and allocation of eye movements to refer to attention (Rayner, 2009).

When it comes to the role of emotional states affecting attention, it is useful to distinguish between an internal affective influence, e.g., the emotional state of the observer, and an external affective influence, e.g., the stimulus valence (Damasio, 1999; Kaspar and König, 2012; Kaspar et al., 2013, 2015; Colombetti, 2014). A situation in which attention is subject to both external and internal affective influences is when one explores web pages of online news portals. On the one hand, such web pages commonly contain positive alongside negative information, whereas, on the other hand, one is in a specific emotional state: Positive, negative, or neutral. Kaspar et al. (2015) used such an environment to investigate the internal and external affective influences in a free-viewing task performed by young adults. The participants' emotional state was primed by a series of either positively or negatively valenced visual stimuli. Subsequently, they had to explore web pages containing both positively and negatively valenced content. An analysis of the eye-tracking data revealed that a negative emotional state marginally elicited a more spatially extensive exploration and that attention for negative news increased in participants who were in a positive emotional state. Thus, the state of the observer and the external affective influence impacted the visual exploration.

The valence of the stimulus influences responses beyond visual exploration. Specifically, approach-behavior is naturally associated with an appraisal of something as "good." In contrast, avoidance behavior is naturally associated with an appraisal of something as "bad." As a consequence, we are faster (Chen and Bargh, 1999) and more accurate (Casasanto and Dijkstra, 2010) when making movements that correspond to their embodied meaning, i.e., approach for good, and avoid for bad. In particular, there is a general bodily tendency to approach positive and avoid negative cues and do so faster than vice versa (Phaf et al., 2014). Moreover, it was shown that positive concepts and percepts are placed close-to-the-body locations, while negative concepts and percepts are placed away-from-the-body locations (Marmolejo-Ramos et al., 2018, 2019). This gives evidence for a general bodily reaction to positive or negative stimuli (Phaf et al., 2014; Sharbanee et al., 2014). For example, spider phobics avoid pictures of spiders more strongly than neutral cues (Rinck and Becker, 2007); socially anxious people avoid smiling and angry faces faster than controls (Heuer et al., 2007); schizophrenic patients with higher levels of oxytocin avoid angry faces faster than controls (Brown et al., 2014); individuals with anorexia-nervosa exhibit a decreased approach bias for food cues (Veenstra and de Jong, 2011); and healthy adults pull positive words faster toward them while pushing negative words faster away (Chen and Bargh, 1999). Thus, the valence of stimuli has a widespread impact on bodily states and actions.

The cognitive mechanisms of the automatic approach bias are still debated. One approach, the concept of embodied cognition, rejects the idea that an agent's cognitive life can be understood without considering the particular morphological, biological, and physiological characteristics of its body (Shapiro, 2011; Engel et al., 2013; Walter, 2014). For instance, language processing (Glenberg and Kaschak, 2002), memory (Casasanto and Dijkstra, 2010), visual-motor recalibration (Bhalla and Proffitt, 1999), or distance estimation (Witt and Proffitt, 2008) all rely on specific body characteristics. Moreover, even our abstract concepts are bodily "grounded" and arise from the body (Barsalou, 2008). That is to say that according to the embodied approach of cognition and affectivity, cognitive and affective phenomena can be fully understood only by taking into account the specific morphological, biological, and physiological details of the agent's body (Shapiro, 2011; Engel et al., 2013; Walter, 2014). In particular, bodily movements are specific to the kind of body we have, and to the environment, we interact with, and are thus naturally meaningful. Similarly, the embodied approach to cognition tries to explain the approach-avoidance behavior (Fridland and Wiers, 2018). Importantly for our study, affective states are also considered within the embodied cognition framework. Stephan et al. (2014) discuss emotions in relation to the body and beyond the body and the brain. Furthermore, Slaby et al. (2016) propose an action-oriented understanding of emotions.

However, although emotional priming has been an important topic in research on top-down influences in overt attention, in particular when it comes to disentangling external and internal affective influences, there is little research using embodied primes (Stoykov et al., 2017). As creatures with specific bodily morphology, our onto- and phylogeny make it natural that positive valence is pulling something toward us while pushing it away is negatively valenced (Fridland and Wiers, 2018). Since the human abdominal region is exceptionally vulnerable, we have to protect it by allowing only trustworthy objects to come close. Since survival requires energy, we have to pull nourishing objects toward us while avoiding rotten, poisonous, unsanitary, or noxious objects. While strangers must typically be kept at bay, procreation, nurturing infants, and giving them love and comfort require social approaching. The idea that approach- and avoidance-behavior is naturally associated with appraisals of something as "good" or "bad" is also in line with embodied accounts of emotions (Niedenthal et al., 2005; Stephan et al., 2014), in particular with Damasio's (Damasio, 2001) "somatic marker" theory, according to which emotions function to direct animals toward what is good and direct them away from what is bad. Hence, these considerations suggest that approaching something or pulling it toward us is naturally meaningful, indicating something is positive.

If the automatic approach bias is indeed a general bodily reaction to positive or negative stimuli (Phaf et al., 2014), we should observe it in healthy adults performing an approach-avoidance task. This type of explanation raises new questions. Namely, if the bodily relation is crucial, we would expect an influence of the stimulus valence and a congruency effect. Body movements that are in line with our preferences (pull toward positive, "good"/push away negative, "bad") should influence our eye movements differently than priming by incongruent actions preferences (pull toward negative, "bad"/push away positive, "good"). Therefore, we aim to answer that question with our design. The congruency effect would give support to the claim

that embodied priming modulates our viewing behavior. That is, here, we are primarily interested in modulating the natural (embodied) action in response to a stimulus e.g., congruent vs. incongruent, as well as investigating the effects on subsequent visual exploratory behavior.

The present study builds on an embodied approach to the automatic approach bias in order to investigate (1) whether healthy adults exhibit a comparable automatic approach bias concerning positively and negatively valenced stimuli and (2) how a positive vs. negative emotional state, induced by a congruent vs. incongruent approach-avoidance task affects their overt attention in a free-viewing task.

METHODS

Participants

Twenty participants (6 male, 18 right-handed, mean age of 22.6 years, standard deviation of 2 years) took part in the experiment. They gave written informed consent before the start of the experiment. Participants received either 9€ or course credits in exchange for their participation. All participants had normal or corrected to normal vision and were not aware of the study's scientific purpose. They were either native German speakers or fluent in the German language. This was important since the presented stimuli included headlines written in German. The ethics committee of Osnabrück University approved the study.

General Apparatus

We presented all stimuli on a 24" LCD monitor (BenQ XL2420T; BenQ, Taipeh, Taiwan) with a refresh rate of 114 Hz. Participants sat 80 cm away from the screen. The experiment was controlled by a PC (Dell) connected to an eye tracker computer via an Ethernet cable. We used a head-mounted eye tracker (Eye Link II system) from SR-Research Ltd. (SR-Research Ltd, Ontario, Canada) to track the participants' eye movements. In turn, the eye tracker was connected to a DOS-based computer (Pentium 4; Dell, Round Rock, TX, USA) running the application software. In total, the eye tracker comprised three infrared cameras. The head camera recorded infrared sensors attached to the monitor's corners to calculate the head position in relation to the screen continuously. This allowed a stable gaze recording irrespective of involuntary small head movements. The other two infrared cameras recorded the participants' pupil positions. The sampling rate of binocular recordings was 500 Hz. The room was darkened during the entire experiment.

A 13-point calibration task preceded each recording. It consisted of fixation points appearing consecutively in random order at various screen locations, and participants were instructed to focus their gaze at these points. Each point had a visual angle size of 0.5°. We validated the calibration by calculating the drift error for each point. Thereby it was assured that the mean validation error stayed below a 0.3° visual angle and the maximum validation error below a 1° visual angle. The calibration was repeated until the mentioned accuracy was reached.

We used the eye tracker's default settings to calculate saccades and fixations. Saccade detection was based on a velocity of at least 30° visual angle/s and acceleration of at least 8,000° visual angles/s^2. To trigger a saccade, the saccade signal had to be sustained for at least 4 ms. By the time the eyes moved significantly from the fixation point (i.e., exceeding a motion threshold), the saccade's temporal and spatial onset had been defined. By default, we set this motion threshold to a 0.1° visual degree. After the saccade onset, the minimal saccade velocity was 25° visual degree/s. Following this, a period without a saccade was marked as fixation. Each trial was followed by a fixation cross appearing in the screen center to control drifts in measurements. The first fixation following each stimulus's onset was excluded from our analysis because this was an artifact from the drift correction before the respective trial's onset.

The joystick used for the approach-avoidance task (Logitech Attack TM 3; Logitech, Apples, Switzerland) was connected to the computer screen. Matlab's Psychtoolbox V3 (r2017a; MathWorks Company) enabled us to record response times (pushing/pulling movements). The joystick was placed on the table in front of the participants. We used MATLAB to preprocess eye-tracking data and R to analyze all data. Analysis scripts and data are available online (https://osf.io/cyz9b/).

Stimuli

The experiment included two separate phases (see below for details). First, participants performed an approach-avoidance task, viewing isolated images. Afterward, participants visually explored web pages, each containing two embedded images with additional text columns in a typical newspaper layout. As we investigated the influence of the approach-avoidance task on later visual exploration, we labeled the isolated images as "primes."

In this study, we used 88 full-colored images from the International Affective Picture Set (IAPS) (Lang et al., 1997). Kaspar et al. (2015) used the identical stimulus set. Half of the images had a valence rated below 3 (IAPS scale) and served as negative primes. The other 44 images had valence ratings above 7 and served as positive primes. To prevent the images from blurring, we presented all of them in their native resolution of 1,024 × 768 pixels on a gray background (RGB values: 182/182/182), centered in the middle of the screen (resolution of 1,920 × 1,080 pixels, **Figure 2A**).

In the present study, twenty-four prototypes of news web pages were used, previously designed by Kaspar et al. (2015) (**Figure 1**). The web page images' resolution fits the screens' resolution (1,920 × 1,080). Two target areas, embedded by several textual and pictorial components, were constructed in one web page design (**Figure 1A**). Each main news article included either a negative or positive IAPS image (615 × 411 pixels), a matching heading, as well as a link to the entire news report. It is important to note that there was no other textual content regarding the main news. This was done to avoid attraction biases because of how appealing news may have been for individuals who participated. Since participants did not interact with the web pages, the link served no function, except for creating a realistic version of a news web page that can be found on the world wide web. The structure and content of the web pages remained the same throughout the whole study. However, the side of the negative and positive content was counterbalanced

FIGURE 1 | Web pages used in the study. **(A)** Prototype web page with different sections: headlines, adds, tabs, news, and images. **(B)** Example of a web page with a negative image on the left side. **(C)** Example of a web page with a negative image on the right side [reverse of **(B)**].

(**Figures 1B,C**). The 48 images, embedded in the 24 web pages, differed from those used in the priming sessions.

The additional elements on each web page were four short news reports about ongoing current affairs. These elements were placed below the main news articles. The frame around the main news articles was completed by flanking advertisements on the left and right sides (**Figure 1A**). Please note that the statistical properties of forward directed saccades and backward directed saccades (regressions) while reading the text do not enter the analysis presented here in any form. As a standard feature on regular web pages, the upper left corner was secured for a tabs region which is necessary for general navigation. Previous work by Kaspar et al. (2015), using the same set of web pages, tested for the possibility that differences in eye movement parameters, within positive and negative images, could evolve from systematic differences in visual saliency. Therefore, a standard algorithm by Itti et al. (1998) that extracts the physical features of images and, based on this, predicts fixation patterns was applied. In addition to this, a graph-based visual saliency (GBVS) developed by Harel et al. (2007) was applied, as it predicts the fixations with a higher probability. After application, no difference regarding the visual saliency was found between the positive and negative images in the stimulus set [both $t_{(35)} \leq 0.941$, $p \geq 0.356$].

Procedure and Design

We divided the participants randomly into two groups. One group started with the congruent block of the approach-avoidance task. The other group started with the incongruent block. In each condition, participants faced a random sequence of 44 images of different valence (22 positive and 22 negative images). As soon as an image was presented, the participants had to respond with the joystick. The task paradigm required participants to push or pull the joystick in response to the image's valence. Participants used their dominant hand to manipulate the joystick in front of them. Participants in the congruent task condition had to pull (approach) the joystick toward themselves whenever a positive image was shown, and push (avoid) the joystick whenever a negative image was shown. In the incongruent condition, participants had to act reversely. They had to pull the negative images toward themselves and push away the positive ones (**Figures 2B,C**). They were instructed to respond as quickly and as accurately as possible. It was not possible to rectify and correct response mistakes. Additionally, while moving the joystick toward or away, the image changed in size. The zoom feature of the approach-avoidance task was programmed in MATLAB's Psychtoolbox V3 (r2017a; MathWorks Company); as such, a shown image smoothly decreased in size as soon as the joystick was pushed (**Figure 2B**). Conversely, the image size increased once the joystick was pulled (**Figure 2C**). It is important to note that participants were instructed to push or pull the joystick to its limit. Overall, participants took about 5 min to complete this first part of the experiment.

In the subsequent eye-tracking session, we recorded the viewing behavior on prototypes of 12 news web pages. Following earlier research of Kaspar et al. (2015) and ensuring the same experimental design, each web page was displayed for 15 s. We instructed participants to explore the web pages freely (free-viewing task).

After the eye-tracking session, participants had a short break. The second part of the experiment, directly after the break, required the participants to complete the joystick approach-avoidance task in the other condition. Participants who performed in the congruent approach-avoidance task had to complete the incongruent condition. The opposite applied to participants in the other group. After the second priming session, an additional 12 web pages were displayed following the same procedure described above (free-viewing task).

RESULTS

Performance in the Embodied Approach-Avoidance Task

For the priming part of the experiment, we first calculated the accuracy of performance to check whether participants followed instructions. In the congruent condition, they had to pull positive and push negative primes. In the incongruent condition, the assigned actions were reversed. We found that participants made a low amount of errors (3.6%). This suggests that instructions were clear, and participants followed them. Therefore, we excluded error trials from any further analysis.

Second, we focused on response times in the experiment's priming part to check whether positive and negative images in congruent and incongruent conditions involve different cognitive processes and, therefore, longer/shorter response times.

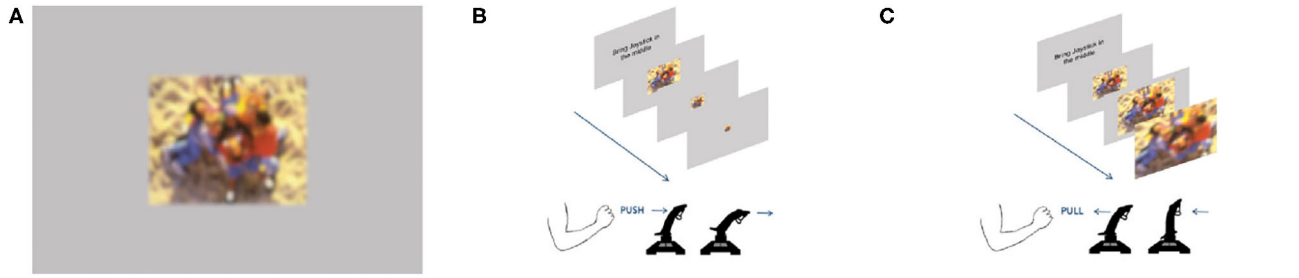

FIGURE 2 | Embodied primes used in the study. **(A)** Example of one priming stimulus: IAPS image with surrounding gray frame. This photograph depicts positive content with children laughing and playing on a sandy playground. Hint: Due to restrictions of showing IAPS images in public, it is masked with a diffusing filter. **(B)** Incongruent condition: joystick pushed forward—positive prime zooms out. **(C)** Congruent condition: joystick pulled backward—positive prime zooms in.

As the data showed a skewed distribution, we log-transformed it before further analysis (we used natural logarithm) (Marmolejo-Ramos et al., 2015). We used a linear mixed model (LMM) to analyze response times. The LMM was calculated with the lme4 package (Bates et al., 2014), and p-values were based on Wald's T-test using the lmerTest package (Kuznetsova et al., 2017). Degrees of freedoms were calculated using the Satterthwaite approximation. We modeled response times by image valence (positive and negative), and experimental condition (congruent and incongruent movements) as fixed effects and interactions between them. As random effects, we used random intercepts for grouping variable participants. For all predictors, we used effect coding scheme with binary factors coded as −0.5 and 0.5. Thus, the resulting estimates can be directly interpreted as the main effects. This coding scheme's advantage is that the fixed effect intercept is estimated as the grand average across all conditions and not a baseline condition average. We found the main effect of the image valence [$t_{(1673.02)} = -3.726$, $p < 0.001$] on response times (**Figure 3B**). The natural logarithm of the response time to negative stimuli was about 0.049 times smaller than to the positive stimuli. This corresponds to a speedup (reduction of response time) by a factor of 5.03%. Furthermore, we found the main effect of the congruency of the task [$t_{(1673.04)} = -4.71$, $p < 0.001$; **Figure 3A**]. The response in the incongruent condition was slower by about 6.39%. The interaction between these effects was not significant [$t_{(1673.02)} = -0.87$, $p > 0.38$]. These results demonstrate independent additive effects of faster movements under congruent conditions and faster movements in response to negative pictures.

Eye Movements in the Free-Viewing Task (Web Pages)

As a next step, we investigated the effect of priming (condition: congruent and incongruent) on the viewing of news pages containing emotional stimuli (valence: negative and positive) on either side (side: left and right). The participants freely viewed different web pages containing one positive image and one negative image and additional filler texts, while we collected eye movement data. We characterized the exploration of these web pages with the two images as regions of interest (ROIs) with a various eye movement measures. Specifically, we used four different measures to quantify eye movements within ROIs: the average fixation duration within each image, the number of fixations within each image, the total dwell time on each image, and the length of saccades within each image. Additionally, we analyzed the number of saccades and their length from the outside to the inside of the images. For all six measures, we used the same statistical procedures. Similarly to the response time analysis, we employed linear mixed models. We modeled each of the variables by experimental condition (congruent and incongruent movements before the free-viewing task), image valence (positive and negative), and side of the image (left and right) as fixed effects and the interactions between them as random effects. We used random intercepts for grouping variable participants. For all predictors, we used effect coding scheme with binary factors coded as −0.5 and 0.5. We visually inspected the normality of the data. All variables, aside from dwell time, were log-transformed to achieve normally distributed data. Jointly, these measures and analyses allow the characterization of viewing behavior on the web pages after priming.

Fixation Duration Within ROIs

As the first measure, we used the average fixation duration within each ROI to measure the depth of processing (Ehinger et al., 2018). We did not find the effect of condition [$t_{(8291.27)} = -0.615$, $p = 0.54$] on fixation duration. However, we found the main effect of the valence [$t_{(8286.74)} = -3.513$, $p < 0.001$; **Figure 4A**]. The average fixation duration on negative images was longer by about 2.8%. Furthermore, we observed the main effect of the side [$t_{(8284.26)} = -3.093$, $p < 0.01$; **Figure 4B**]. Fixations on the image displayed on the left side were longer by about 2.45%. Further, we found the significant interaction between valence and side [$t_{(8283.33)} = -3.318$, $p < 0.001$, **Figure 4C**]. The difference in fixation duration on positive and negative images was larger on the right side. The size of this interaction was of the same order of magnitude as the main effect of the side of the image. That is, the longest average fixation duration was observed for the combination of negative images on the right side of the displayed web page. All other two-way and three-way interactions were not significant ($p > 0.18$). These results show that the displayed web page parameters, i.e., valence and side, had a significant influence on the depth of processing at

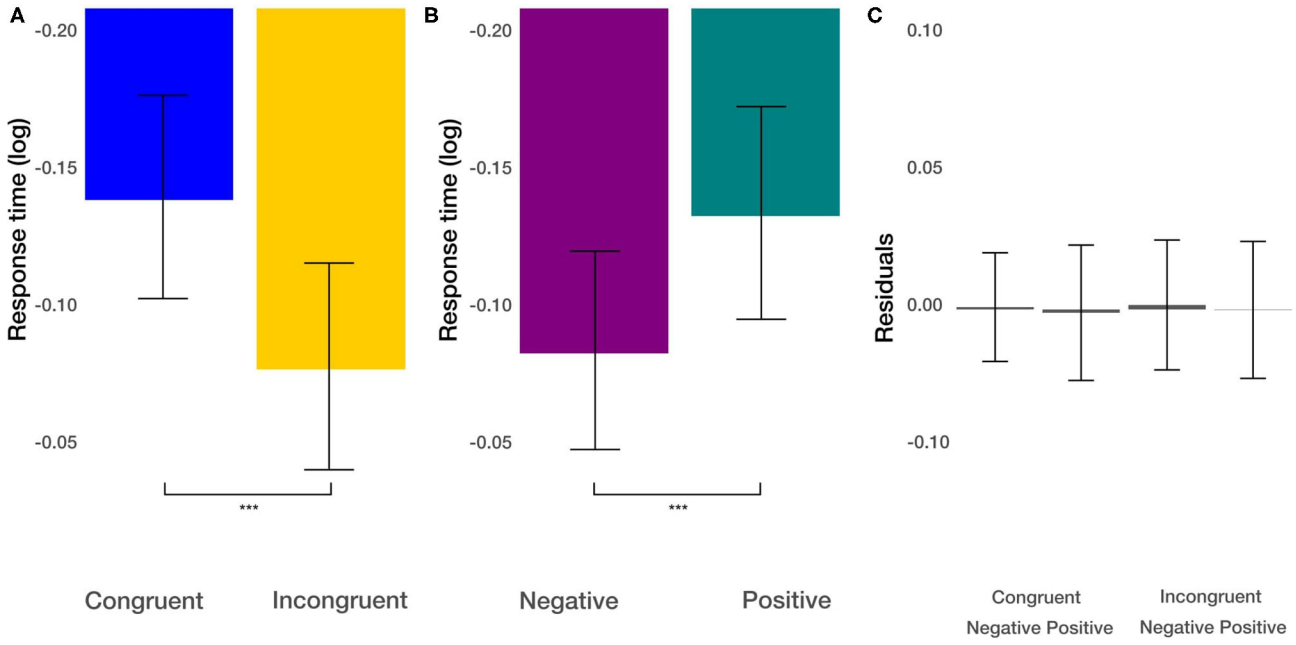

FIGURE 3 | Response time results. **(A)** The main effect of the congruency of the priming. **(B)** The main effect of image valence (negative vs. positive). **(C)** Averaged residuals of all condition combinations from response time linear mixed model. In all plots, bars represent mean values and whiskers standard error of the mean. Stars represent significance levels: ***$p < 0.001$.

individual fixation locations. However, this did not apply to the priming condition.

Number of Fixations Within ROIs

Next, we considered the number of fixations within each image to measure attention devoted to the respective stimulus. We found the main effect of condition [$t_{(915.2)} = -2.271$, $p = 0.0234$; **Figure 5A**]. The number of fixations within the ROIs after congruent priming was larger by 11.47%. Furthermore, we observed the main effect of the valence [$t_{(915.19)} = -5.112$, $p < 0.001$; **Figure 5B**]. Negative images captured 27.68% more fixations than positive images. Finally, we found the main effect of the side [$t_{(915.17)} = -2.816$, $p < 0.01$; **Figure 5C**]. Images displayed on the left side captured 14.41% more fixations. All two-way and three-way interactions were not significant ($p > 0.1$). These results demonstrate additive effects, in terms of the logarithm of the number of fixations. Converted back to the number of fixations within the ROIs, this results in multiplicative effects on the condition, valence, and side on the attention devoted to the images.

Dwell Time Within ROIs

The dwell time combines the aspects of fixation duration and the number of fixations within the ROIs. We found the main effect of condition [$t_{(915.2)} = -2.274$, $p = 0.0232$; **Figure 6A**]. The dwell time within the ROIs after congruent priming was on average 190 ms larger. Furthermore, we observed the main effect of the valence [$t_{(915.2)} = -6.146$, $p < 0.001$; **Figure 6B**]. Dwell time on negative images was on average 513 ms larger than on positive images. Finally, we observed the main effect of the side [$t_{(915.17)} = -3.381$, $p < 0.01$; **Figure 6C**]. On average, the dwell time within images was on average 282 ms larger on the left side. All two-way and three-way interactions were not significant ($p > 0.33$). These results resemble the results in the analysis of the number of fixations within ROIs. They provide evidence for independent effects of the priming condition, the valence of the viewed image, and the side of image location on the dwell time.

Saccade's Length Within ROIs

As a measure of exploration within the images, we used the saccadic length. We did not find the main effect of condition on the saccadic length [$t_{(8296.9)} = 1.321$, $p = 0.187$]. However, we did find the main effect of the image valence [$t_{(8291.31)} = 4.253$, $p < 0.001$; **Figure 7A**]. Within negative images, saccades were shorter by 9.55%. Further, we observed a small but significant main effect of the side [$t_{(8287.02)} = 2.618$, $p < 0.01$; **Figure 7B**] on the saccade's length. Saccades were shorter by 5.76% on the left side. Furthermore, we found significant two-way interactions between the image valence and the side of the image [$t_{(8285.35)} = -2.038$, $p = 0.0415$; **Figure 7C**], with a slightly larger difference in the saccadic length for positive and negative images on the left side. Additionally, we observed the interaction of condition and side [$t_{(8289.3)} = 2.907$, $p < 0.01$; **Figure 7D**]. Whereas, images displayed on the left condition were trivial, images on the right side were explored by longer saccades after incongruent priming. We did not find the interaction between condition and valance ($p > 0.96$), as well as no three-way interaction between all factors ($p > 0.27$). These results give evidence for a more focused

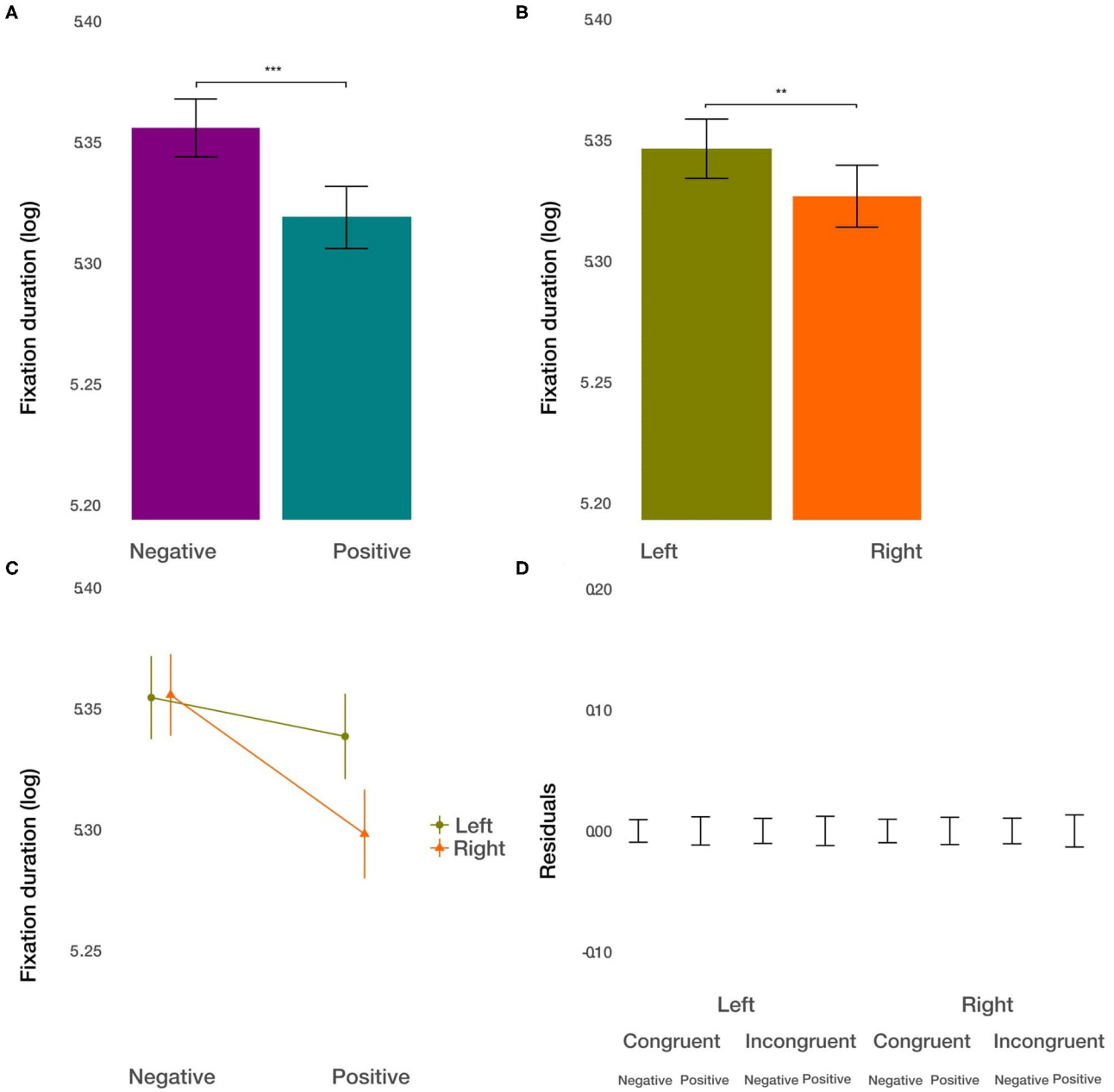

FIGURE 4 | Fixation duration results. **(A)** The main effect of image valence (negative vs. positive). **(B)** The main effect of the side (left vs. right). **(C)** The interaction between valence and side. **(D)** Averaged residuals of all condition combinations from the fixation duration linear mixed model. In all plots, bars represent mean values and whiskers standard error of the mean. Stars represent significance levels: $^{**}p < 0.01$ and $^{***}p < 0.001$.

exploration of images with negative valence, specifically when displayed on the left side. The priming condition modulated the influence of the side with a larger differential effect on the exploration of images displayed on the right side.

Number and Length of Saccades From Outside ROIs Toward ROIs

As a measure of how well the images can attract attention, we utilized the number of saccades from outside the image toward the inside. We found a trend toward significance for the main effect of the congruency of the priming [$t_{(850.76)} = -1.894$, $p = 0.0586$]. Participants, on average, made 6.96% more saccades from outside into an image after congruent condition. The main effect of the valence of the image [$t_{(851.31)} = -1.936$, $p = 0.0532$] on the number of saccades from outside images toward them missed the significance threshold. Nominally, participants made 7.13% more saccades on negative images. The effect of the side of the image and all two-way

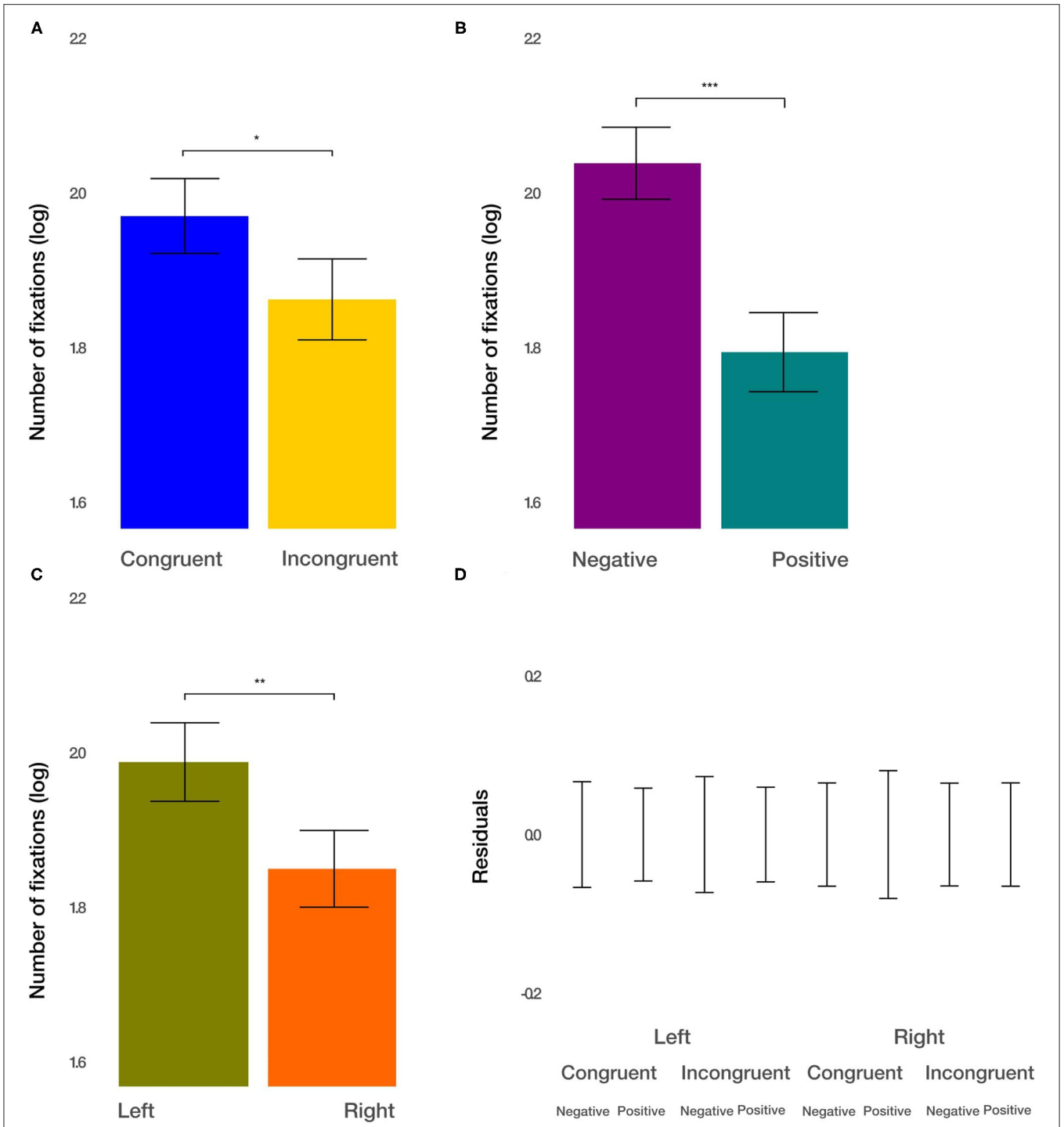

FIGURE 5 | The number of fixations results. **(A)** The main effect of the congruency of the priming. **(B)** The main effect of valence (negative vs. positive). **(C)** The main effect of the side (left vs. right). **(D)** Averaged residuals of all condition combinations from the number of fixations linear mixed model. In all plots, bars represent mean values and whiskers standard error of the mean. Stars represent significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

and three-way interactions were not significant (all $p > 0.12$). Furthermore, we analyzed the length of saccades from outside images toward them. We found the main effect of congruency of the priming [$t_{(2011.69)} = 2.189$, $p = 0.0287$]. After priming with congruent actions, the average saccade length from outside into the images was shorter by 9.71%. Furthermore, the main effect of the side of the image [$t_{(2009.88)} = 2.269$, $p = 0.0234$) on the saccade's length from outside images toward them was

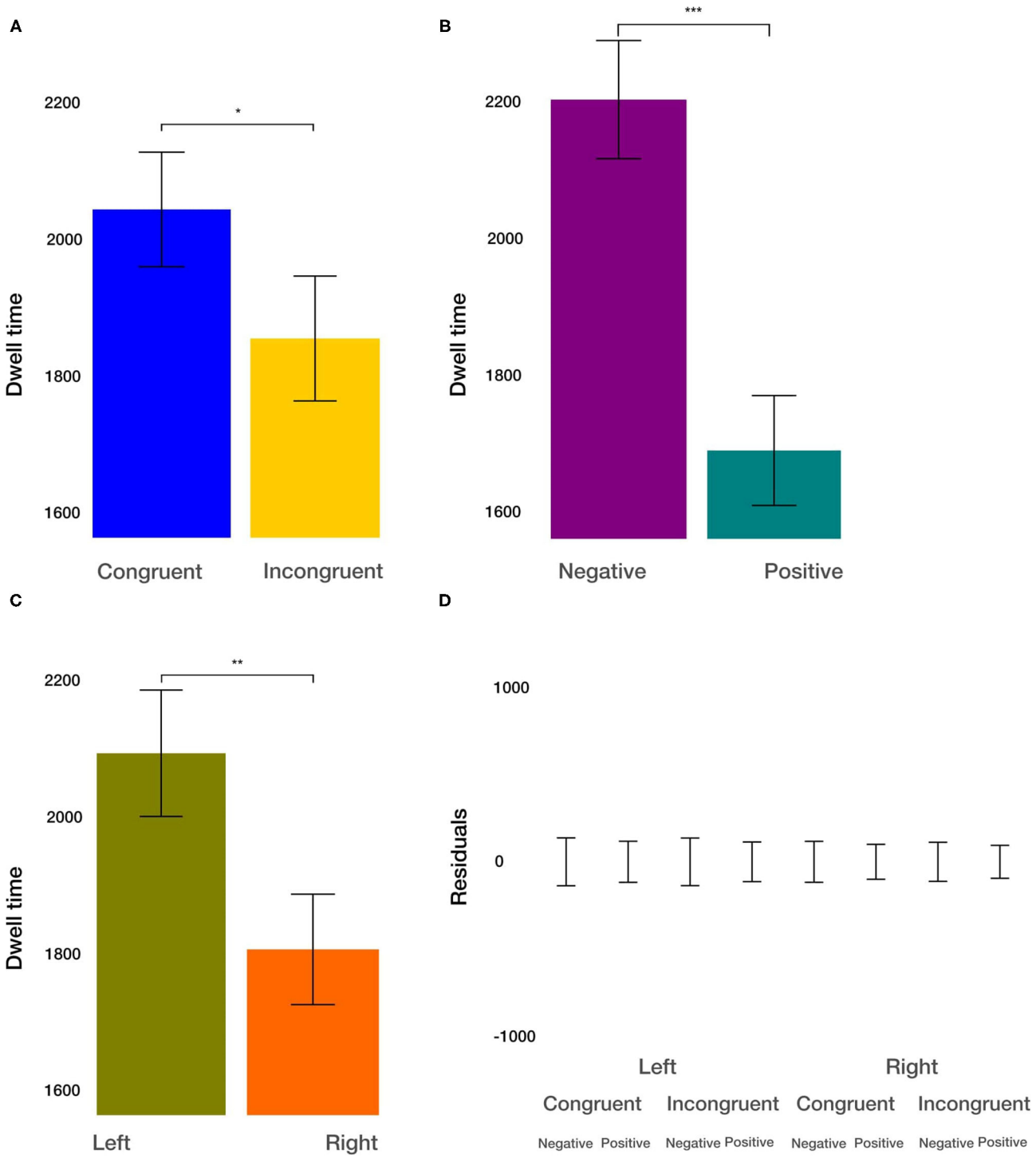

FIGURE 6 | Dwell time results. **(A)** The main effect of the congruency of the priming. **(B)** The main effect of image valence (negative vs. positive). **(C)** The main effect of the side (left vs. right). **(D)** Averaged residuals of all condition combinations from the dwell time linear mixed model. In all plots, bars represent mean values and whiskers standard error of the mean. Stars represent significance levels: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

significant. Saccades targeting the left image were, on average, shorter by 10.07%. The effect of the image valence and all two-way and three-way interactions were not significant (all $p > 0.06$). These results suggest that participants, on average, make shorter and more saccades after congruent priming from outside images toward them. Furthermore, on average, participants make longer saccades from outside images toward the images on the right side.

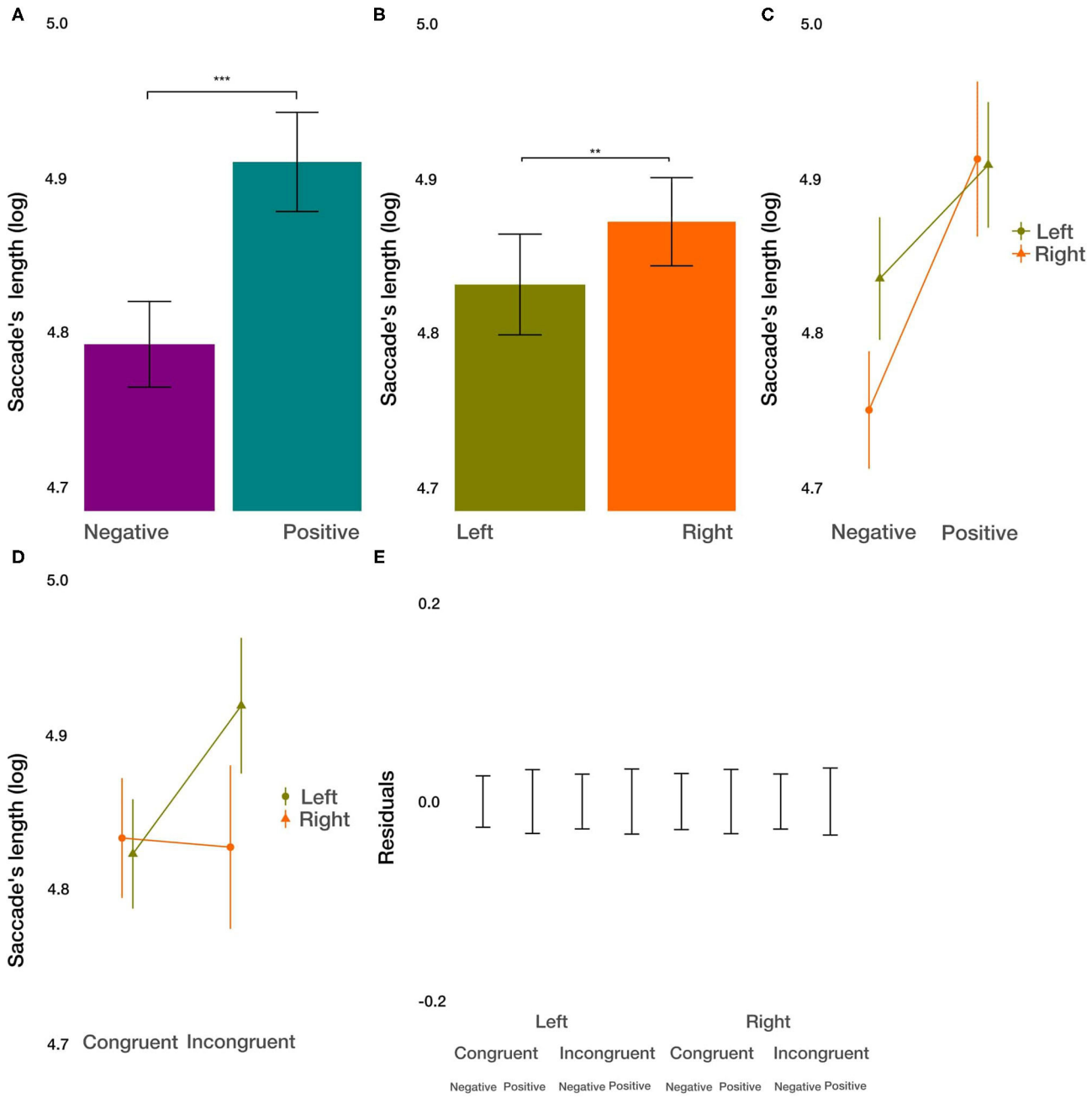

FIGURE 7 | Saccade's length results. **(A)** The main effect of image valence (negative vs. positive). **(B)** The main effect of the side (left vs. right). **(C)** The interaction between image valence and side of the image. **(D)** The interaction between congruency of the priming and side of the image. **(E)** Averaged residuals of all condition combinations from the saccade's length linear mixed model. In all plots, bars represent mean values and whiskers standard error of the mean. Stars represent significance levels: $^{**}p < 0.01$ and $^{***}p < 0.001$.

DISCUSSION

In the present study, we used an active bodily interaction with affective stimuli in an approach-avoidance task to investigate the influence on the later free visual exploration of news web pages containing emotional images. First, positively or negatively valenced images were zoomed in or out by pulling or pushing the joystick. Here we found multiplicative effects of valence and condition. That is, we could replicate the results of previous studies and report a faster response in the congruent condition. Furthermore, negative stimuli were reacted to faster. The lack of interaction and the additive effects on the log-response time suggest independent multiplicative effects on the base response time. Second, concerning the influence of embodied

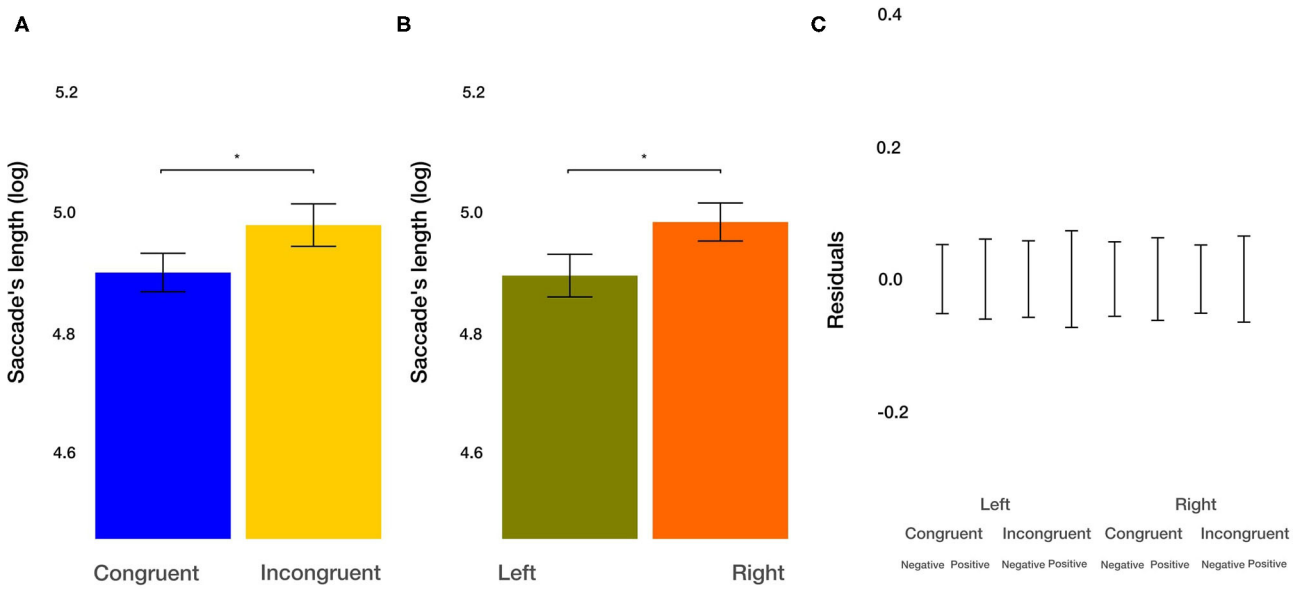

FIGURE 8 | Saccade's length from outside ROIs toward ROIs results. **(A)** The main effect of the congruency of the priming. **(B)** The main effect of the side of the image (left vs. right). **(C)** Averaged residuals of all condition combinations from the saccade's length from outside ROIs toward ROIs linear mixed model. In all plots, bars represent mean values and whiskers standard error of the mean. Stars represent significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

priming on eye movements in the subsequent free-viewing task, we observed the main effects of valence, side of the presentation, condition, as well as specific interactions. This demonstrates the influence of stimulus properties (valence), internal variables (priming by condition), and spatial properties (side) on visual exploration.

The following discussion will address the two main parts of our analyses. We will first discuss the viewing behavior (fixations and saccades) made only in the emotion-laden main news. The subsequent part deals with the approach and avoidance behavior while performing the approach-avoidance task while taking the response times into account.

First, we observed an influence of the priming condition, i.e., performing congruent or incongruent actions in the approach-avoidance task on later visual exploration. Specifically, after priming in the congruent condition, the number of fixations on images increased, the total dwell time increased, and the average saccade length from outside of the images toward the images decreased. The combination of these effects suggests increased attention to web pages' image content after the subjects performed congruent actions on images in the priming phase. Thus, congruent bodily interaction with images in the priming phase fosters visual interaction in the subsequent exploration phase.

Second, we found systematic effects of the valence of images in the exploration phase. Specifically, on images with negative valence, the average fixation duration was prolonged, more fixations were completed, and the total dwell time increased. Particularly, the average length of saccades within the images of negative valence decreased. This combination of effects speaks for increased scrutiny of negative valenced images.

Third, we observed a lateral asymmetry in the visual exploration phase. On average, participants displayed longer fixations, more fixations, and longer dwell time on the left side. The change of the length of saccades within the images was significant but quantitatively not relevant. In contrast, the saccades' length from the outside of the images toward the inside was shorter on the left side. The combination of these effects largely resembles the effects observed with respect to the priming condition and suggests increased visual interaction with stimuli presented on the left side. These results further corroborate results suggesting spatial biases in eye movements (Ossandon et al., 2014). This suggests that biases toward the left side influence not only the number of fixations but the major properties of visual exploration as well.

Finally, with respect to interactions of condition, valence, and side, it is noteworthy that there were only a few. For the number of fixations, dwell time, and the average length of saccades from outside to inside, we did not observe any 2-way or 3-way interactions, and the residuals after discounting for the main effects were relatively small (**Figures 3C**, **4D**, **5D**, **6D**, **7E**, **8**). Only for the average saccadic length within the images did we observe an interaction of valence*side and condition*side and for the average fixation duration an interaction of valence*side. It appears that with respect to the saccadic length within the images for the left side, the valence is more important, and the priming condition less important for images on the left side. The fixation duration is less affected by image valence on the left side. Overall, it is striking that the effects of the three independent variables are largely independent, and the interactions are limited to a few aspects.

The results of this study suggest that approach and avoidance reactions in humans have a direct influence on attention

allocation and gaze behavior. We used the embodied cognition approach and, more specifically, the approach-avoidance task to explore its effect on eye movements. This study adds to the limited amount of eye-tracking research that has dealt with the interplay of top-down influences and bottom-up features.

To induce a positive or a negative emotional state, Kaspar et al. (2015) had participants watch either positive or negative sequences of 44 full-colored images from the International Affective Picture System (IAPS; Lang et al., 1997) with a valence rating below 3 for negative and a valence rating above 7 for positive primes. In the subsequent eye-tracking session, they presented 24 similarly structured webpages that included a positive and a negative IAPS image: one on the left, the other on the right. They found that a negative emotional state marginally elicited a more spatially extensive exploration. In our study, we used the same news web pages. However, instead of inducing emotional states by passively watching pictures, we used an approach-avoidance task as an embodied prime for positive and negative emotional states. In contrast, no specific emotional valence was primed, but rather the congruent or incongruent action, i.e., approach/avoidance of positive/negative valenced stimuli or the reverse assignment. There is ample evidence that our emotions affect our visual behavior. Regarding the direct effect of emotions on visual exploration, the broaden-and-build model of positive emotions (Fredrickson, 1998) claims that positive emotions such as joy, interest, elation, or love, temporarily expand the focus of attention, therefore, increasing the thought-action repertoire by fostering interest in the environment and encouraging play and exploration (Fredrickson, 2000). Accordingly, being in a happy emotional state vs. being in a sad or neutral emotional state has been shown to increase participants' breadth of attention (Rowe et al., 2007). Similarly, Wadlinger and Isaacowitz (2006) found that the distribution of participants' fixations on an image is broader in individuals induced into a positive emotional state, with more frequent saccades to neutral or positively valenced parts and with more fixations on positively valenced peripheral stimuli. Whereas, broadened attention is often associated with anxiety (Gruzelier and Phelan, 1991), which has led some to speculate that this might be an adaption to a negative emotional state (Garland et al., 2011), while a positive emotional state may reduce the motivation to scrutinize the environment because of an increased feeling of safety (Schwarz, 1990). Part of the explanation for these diverse findings may be that the emotional state induction procedures are also diverse, particularly concerning neutral emotional states. For instance, whereas some actively induce a positive emotional state by offering participants a bag of candies but simply do nothing in the neutral condition (Wadlinger and Isaacowitz, 2006), others rely on a waiting room manipulation to actively induce a neutral emotional state (Herz et al., 2004). Others have also been known to use movies (Grubert et al., 2013) or music (Shapiro and Lim, 1989). According to Kaspar et al. (2015), this diversity of emotional state induction method questions the assumption that a neutral emotional state is always an adequate control condition. This may help explain why being in a negative emotional state had the same effect as being in a neutral emotional state according to some studies (Rowe et al., 2007). Whereas, other studies found similar effects of being in a positive and neutral emotional state (Chipchase and Chapman, 2013). In light of this still unresolved issue, the present study followed (Kaspar et al., 2015) and solely contrasted positive with negative emotional states and focused on the effects of priming in congruent vs. incongruent actions in an approach-avoidance task.

When investigating embodied cognition, a high degree of ecological validity is necessary. We instructed participants to use a joystick to either approach or avoid positively or negatively valenced pictures displayed on a screen (Ernst et al., 2013). To increase immersion, we implemented a visual "zooming-effect" (Rinck and Becker, 2007). When the joystick was pushed, the images were zoomed out, and when it was pulled, the images were zoomed in. This not only ensured a more realistic impression of movement toward or away from the images, but it also illustrated any ambiguity in the participants' arm movements. The appraisal of a movement depends upon what is achieved (Lavender and Hommel, 2007; Krieglmeyer et al., 2010). Stretching out one's arm often indicates a negatively valenced avoidance-behavior, i.e., when a harmful object is pushed away. Yet, it can also be an indispensable part of a positively valenced approach-behavior, for instance when one reaches for nourishing food or one's infant. A joystick-based Approach avoidance task with a zooming-effect resolves this ambivalence. To further increase the immersion utilizing techniques of virtual reality offer themselves.

In addition to the effect of participants' emotional states on their attention, we also explored the approach and avoidance behavior in the priming conditions. Since the IAPS pictures have exhibited an impact on the emotions of participants and therefore serve as a reliable priming method (Kaspar et al., 2015), we made use of them in our study to also modulate congruent vs. incongruent actions by the participants. Previous study designs have solely presented participants with a row of pictures within one category. However, our study design differs from previous work in that we let the participants visually and physically interact with the depicted pictures. For this reason, we joined pictures of two valence categories in one task, which had to be treated differently. Since we were working with IAPS images, it is worthy of mentioning that the highly negative images were also accompanied by a higher level of arousal, in contrast to highly positive images (Lang et al., 1997). This applies to the IAPS images within the priming block and the images embedded in the news web pages. As Kaspar et al. (2015) note, negative emotions, such as anxiety, anger, and fear, also happen to be more arousing for the participants compared to positive emotions, such as pride or happiness. This applies as well to negative and positive emotional conditions. As mentioned in the introduction, along with the increment of arousal in negative emotions comes an increase in attention. This is caused by the initiation of survival-related actions related to behavioral and physical fight-or-flight responses (Fredrickson, 2000). In turn, the specific arousal, which is immediately elicited by the mere presence of the valenced images, is interwoven with the arousal that is elicited by the interactive treatment of the images in the priming condition. Since the primes used in this study comprised of both valence categories, it is challenging to make any explicit distinction.

However, the approach-avoidance task served as an authentic method to strive for and impact the internal approach and avoidance reactions in humans. Participants in the incongruent priming condition were significantly slower in treating the primes as instructed. Thus, they were slower to pull negative images toward themselves and push positive images away from themselves. The difficulty of the incongruent priming condition task was reflected in the participants' response times. In general, the task instruction (to pull negative images toward oneself) essentially acts against the avoidance effect, which has been presented as an example of embodied cognition that emphasizes action-oriented behavior, i.e., actions related to survival. A direct comparison of both task conditions clearly revealed the avoidance effect. In the congruent condition, participants were significantly faster to avoid the negative stimuli, compared with avoiding the positive stimuli in the incongruent condition.

In controlled attentional shifts, older adults show a positivity bias and negativity avoidance (Isaacowitz et al., 2006). In contrast, no such bias is observable for automatic attentional shifts (Hahn et al., 2006; Mather and Knight, 2006; Knight et al., 2007). The results are inconclusive for younger adults. Some studies find a preference for negative stimuli (Thomas and Hasher, 2006; Tomaszczyk et al., 2008), while others report a tendency to avoid negative stimuli (Becker and Detweiler-Bedell, 2009). Some studies find emotional state-incongruent preferences (Parrott and Sabini, 1990; Schwager and Rothermund, 2013), while others report emotional state-congruent preferences (Ferraro et al., 2003; Isaacowitz et al., 2008; Koster et al., 2010; Becker and Leinenger, 2011). Presumably, this inconsistency is partly because studies only focused on external affective influences and disregarded the participants' emotional state. However, considering the participants' emotional states is crucial because one's emotional state can determine one's current goals. In fact, when emotional regulation is the primary goal of younger adults, they focus less on negative images and more on positive images (Xing and Isaacowitz, 2006). Moreover, students who learn to focus on positive stimuli subsequently show reduced attention for negative stimuli (Wadlinger and Isaacowitz, 2006), indicating that attention is a powerful tool for emotional states regulation (Wadlinger and Isaacowitz, 2011). In contrast, Das and Fennis (2008) found that a positive emotional state can increase attention for negative information. However, the primary goal of young adults exploring news pages is arguably not emotional state regulation. They are rather in a "browsing mode" in which they search for personally interesting information. In such a mode, features of the stimulus, such as its valence, are more likely to catch the observer's attention (Hamborg et al., 2012). In contrast to these mixed results, the effects of congruent vs. incongruent conditions in the present study are relatively straightforward.

Many studies investigated an automatic approach bias in patients with substance abuse disorders. Individuals with a substance abuse disorder exhibit an automatic bias toward drug-related words (Cox et al., 2006) or pictures (Field et al., 2013). In stimulus-response compatibility tasks, in which participants have to use a joystick to move cues either away or toward themselves, they approach rather than avoid drug-related cues and they approach them faster than they avoid them. In an implicit approach-avoidance task, in which participants push and pull cues according to formal features [like the format of a picture (Wiers et al., 2011) or its vertical alignment (Cousijn et al., 2011)], heavy drinkers (Wiers et al., 2009), patients with alcohol abuse disorder (Wiers et al., 2011, 2014), heroin addicts (Zhou et al., 2012), smokers (Wiers C.E. et al., 2013; Wiers R.W. et al., 2013a,b), and cannabis users (Cousijn et al., 2011) approach drug cues faster than healthy controls. In an explicit approach-avoidance task in which participants either push away drug cues while pulling neutral cues toward them or vice versa, individuals with alcohol abuse disorder approach drug cues faster than they avoid them (Ernst et al., 2014). Thus, there is accumulating evidence for a general automatic approach/avoidance bias related to substance abuse.

Essentially, with our empirical work we cannot address the dispute on causal and constitutive relationships (Kaiser and Krickel, 2017). In the spirit of hypothesis testing we present data, that are compatible with the framework of embodied cognition. Our results could be further explained by two different aspects (limitations) that we address here. First, one of the limitations is that we do not know whether embodied priming with emotionally laden pictures triggered any emotions. One could test that by measuring different levels of arousal during the priming phase. However, we did not investigate the exact physiological basis behind embodied priming but its influence on the viewing behavior. In the future, it would be worth also studying these physiological underpinnings of embodied priming. Second, we used binary categories (positive and negative images) in our study based on the validated dataset (Lang et al., 1997). Emotions and perception of emotionally laden images can vary between participants, and therefore, one could additionally improve understanding how the effect that we found emerged if participants rated the imaged themselves. These issues do not change our interpretation, but it is essential to consider them.

In summary, we present how congruent embodied priming influences eye movements in a free-viewing task. Results presented in our study suggest that prior congruent movements, in line with our bodily reactions, can influence how we scrutinize images presented on the World Wide Web.

We found that movements in line with our bodily reactions (approach positive and avoid negative) influence how we observe images presented on the World Wide Web.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of the Osnabrück University. The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PK and SW: conceived the study. FA and PK: study design. FA: data collection. AC, FA, and PK: data analysis and revisions and finalizing the manuscript. AC: initial draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

We gratefully acknowledge the support by the DFG-funded Research Training Group Situated Cognition (GRK 2185/1), Niedersächsischen Innovationsförderprogramms für Forschung und Entwicklung in Unternehmen (NBank)—EyeTrax, and the Deutsche Forschungsgemeinschaft (DFG) Open Access Publishing Fund of Osnabrück University. Moreover, we would like to thank Clayton Thompson for feedback on the manuscript.

REFERENCES

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv:1406.5823 [stat]*. doi: 10.18637/jss.v067.i01

Becker, M. W., and Detweiler-Bedell, B. (2009). Short article: early detection and avoidance of threatening faces during passive viewing. *Q. J. Exp. Psychol.* 62, 1257–1264. doi: 10.1080/17470210902725753

Becker, M. W., and Leinenger, M. (2011). Attentional selection is biased toward mood-congruent stimuli. *Emotion* 11, 1248–1254. doi: 10.1037/a0023524

Bhalla, M., and Proffitt, D. R. (1999). Visual–motor recalibration in geographical slant perception. *J. Exp. Psychol.* 25:1076. doi: 10.1037/0096-1523.25.4.1076

Brown, E. C., Tas, C., Kuzu, D., Esen-Danaci, A., Roelofs, K., and Brüne, M. (2014). Social approach and avoidance behaviour for negative emotions is modulated by endogenous oxytocin and paranoia in schizophrenia. *Psychiatry Res.* 219, 436–442. doi: 10.1016/j.psychres.2014.06.038

Casasanto, D., and Dijkstra, K. (2010). Motor action and emotional memory. *Cognition* 115, 179–185. doi: 10.1016/j.cognition.2009.11.002

Chen, M., and Bargh, J. A. (1999). Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. *Pers. Soc. Psychol. Bull.* 25, 215–224. doi: 10.1177/0146167299025002007

Chipchase, S. Y., and Chapman, P. (2013). Trade-offs in visual attention and the enhancement of memory specificity for positive and negative emotional stimuli. *Q. J. Exp. Psychol.* 66, 277–298. doi: 10.1080/17470218.2012.707664

Colombetti, G. (2014). *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262019958.001.0001

Cousijn, J., Goudriaan, A. E., and Wiers, R. W. (2011). Reaching out towards cannabis: approach-bias in heavy cannabis users predicts changes in cannabis use: approach-bias and cannabis use. *Addiction* 106, 1667–1674. doi: 10.1111/j.1360-0443.2011.03475.x

Cox, W. M., Fadardi, J. S., and Pothos, E. M. (2006). The addiction-stroop test: theoretical considerations and procedural recommendations. *Psychol. Bull.* 132, 443–476. doi: 10.1037/0033-2909.132.3.443

Damasio, A. (2001). Fundamental feelings. *Nature* 413, 781–781. doi: 10.1038/35101669

Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Houghton Mifflin Harcourt.

Das, E., and Fennis, B. M. (2008). In the mood to face the facts: when a positive mood promotes systematic processing of self-threatening information. *Motiv. Emot.* 32, 221–230. doi: 10.1007/s11031-008-9093-1

Ehinger, B. V., Kaufhold, L., and König, P. (2018). Probing the temporal dynamics of the exploration–exploitation dilemma of eye movements. *J. Vis.* 18:6. doi: 10.1167/18.3.6

Einhäuser, W., Spain, M., and Perona, P. (2008). Objects predict fixations better than early saliency. *J. Vis.* 8:18. doi: 10.1167/8.14.18

Einhäuser, Wolfgang, U. R. and Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *J. Vis.* 8:2. doi: 10.1167/8.2.2

Engel, A. K., Maye, A., Kurthen, M., and König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends Cogn. Sci.* 17, 202–209. doi: 10.1016/j.tics.2013.03.006

Ernst, L. H., Plichta, M. M., Dresler, T., Zesewitz, A. K., Tupak, S. V., Haeussinger, F. B., et al. (2014). Prefrontal correlates of approach preferences for alcohol stimuli in alcohol dependence: approach bias for alcohol. *Addict. Biol.* 19, 497–508. doi: 10.1111/adb.12005

Ernst, L. H., Plichta, M. M., Lutz, E., Zesewitz, A. K., Tupak, S. V., Dresler, T., et al. (2013). Prefrontal activation patterns of automatic and regulated approach-avoidance reactions-A functional near-infrared spectroscopy (fNIRS) study. *Cortex* 49, 131–142. doi: 10.1016/j.cortex.2011.09.013

Ferraro, F. R., King, B., Ronning, B., Pekarski, K., and Risan, J. (2003). Effects of induced emotional state on lexical processing in younger and older adults. *J. Psychol.* 137, 262–272. doi: 10.1080/00223980309600613

Field, M., Mogg, K., Mann, B., Bennett, G. A., and Bradley, B. P. (2013). Attentional biases in abstinent alcoholics and their association with craving. *Psychol. Addict. Behav.* 27, 71–80. doi: 10.1037/a0029626

Fredrickson, B. L. (1998). What good are positive emotions? *Rev. Gen. Psychol.* 2, 300–319. doi: 10.1037/1089-2680.2.3.300

Fredrickson, B. L. (2000). Cultivating positive emotions to optimize health and well-being. *Prevent. Treat.* 3:1a. doi: 10.1037/1522-3736.3.1.31a

Fridland, E., and Wiers, C. E. (2018). Addiction and embodiment. *Phenomenol. Cogn. Sci.* 17, 15–42. doi: 10.1007/s11097-017-9508-0

Garland, E. L., Gaylord, S. A., and Fredrickson, B. L. (2011). Positive reappraisal mediates the stress-reductive effects of mindfulness: an upward spiral process. *Mindfulness* 2, 59–67. doi: 10.1007/s12671-011-0043-8

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565. doi: 10.3758/BF03196313

Grubert, A., Schmid, P., and Krummenacher, J. (2013). Happy with a difference, unhappy with an identity: observers' mood determines processing depth in visual search. *Attent. Percept. Psychophys.* 75, 41–52. doi: 10.3758/s13414-012-0385-x

Gruzelier, J., and Phelan, M. (1991). Stress induced reversal of a lexical divided visual-field asymmetry accompanied by retarded electrodermal habituation. *Int. J. Psychophysiol.* 11, 269–276. doi: 10.1016/0167-8760(91)90021-O

Hahn, S., Carlson, C., Singer, S., and Gronlund, S. D. (2006). Aging and visual search: automatic and controlled attentional bias to threat faces. *Acta Psychol.* 123, 312–336. doi: 10.1016/j.actpsy.2006.01.008

Hamborg, K.-C., Bruns, M., Ollermann, F., and Kaspar, K. (2012). The effect of banner animation on fixation behavior and recall performance in search tasks. *Comput. Hum. Behav.* 28, 576–582. doi: 10.1016/j.chb.2011.11.003

Harel, J., Koch, C., and Perona, P. (2007). "Graph-based visual saliency," in *Advances in Neural Information Processing Systems 19 (NIPS 2006)* (Cambridge, MA: MIT Press), 545–552.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., and Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *J. Vis.* 3:6. doi: 10.1167/3.1.6

Herz, R. S., Schankler, C., and Beland, S. (2004). Olfaction, emotion and associative learning: effects on motivated behavior. *Motiv. Emot.* 28, 363–383. doi: 10.1007/s11031-004-2389-x

Heuer, K., Rinck, M., and Becker, E. S. (2007). Avoidance of emotional facial expressions in social anxiety: the approach-avoidance task. *Behav. Res. Ther.* 45, 2990–3001. doi: 10.1016/j.brat.2007.08.010

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., and Welsh, T. N. (2019). No one knows what attention is. *Attent. Percept. Psychophys.* 81, 2288–2303. doi: 10.3758/s13414-019-01846-w

Isaacowitz, D. M., Toner, K., Goren, D., and Wilson, H. R. (2008). Looking while unhappy: mood-congruent gaze in young adults, positive gaze in older adults. *Psychol. Sci.* 19, 848–853. doi: 10.1111/j.1467-9280.2008.02167.x

Isaacowitz, D. M., Wadlinger, H. A., Goren, D., and Wilson, H. R. (2006). Is there an age-related positivity effect in visual attention? A comparison of two methodologies. *Emotion* 6, 511–516. doi: 10.1037/1528-3542.6.3.511

Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res.* 40, 1489–1506. doi: 10.1016/S0042-6989(99)00163-7

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259. doi: 10.1109/34.730558

Kaiser, M. I., and Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *Br. J. Philos. Sci.* 68, 745–779. doi: 10.1093/bjps/axv058

Kaspar, K., Gameiro, R. R., and König, P. (2015). Feeling good, searching the bad: positive priming increases attention and memory for negative stimuli on webpages. *Comput. Hum. Behav.* 53, 332–343. doi: 10.1016/j.chb.2015.07.020

Kaspar, K., Hloucal, T.-M., Kriz, J., Canzler, S., Gameiro, R. R., Krapp, V., et al. (2013). Emotions' impact on viewing behavior under natural conditions. *PLoS ONE* 8:e52737. doi: 10.1371/journal.pone.0052737

Kaspar, K., and König, P. (2012). Emotions and personality traits as high-level factors in visual attention: a review. *Front. Hum. Neurosci.* 6:321. doi: 10.3389/fnhum.2012.00321

Knight, M., Seymour, T. L., Gaunt, J. T., Baker, C., Nesmith, K., and Mather, M. (2007). Aging and goal-directed emotional attention: distraction reverses emotional biases. *Emotion* 7, 705–714. doi: 10.1037/1528-3542.7.4.705

Kollmorgen, S., Nortmann, N., Schröder, S., and König, P. (2010). Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. *PLoS Comput. Biol.* 6, e1000791. doi: 10.1371/journal.pcbi.1000791

Koster, E. H., De Raedt, R., Leyman, L., and De Lissnyder, E. (2010). Mood-congruent attention and memory bias in dysphoria: exploring the coherence among information-processing biases. *Behav. Res. Ther.* 48, 219–225. doi: 10.1016/j.brat.2009.11.004

Krieglmeyer, R., Deutsch, R., De Houwer, J., and De Raedt, R. (2010). Being moved: valence activates approach-avoidance behavior independently of evaluation and approach-avoidance intentions. *Psychol. Sci.* 21, 607–613. doi: 10.1177/0956797610365131

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). *International Affective Picture System (IAPS): Technical Manual and Affective Ratings*. NIMH Center for the Study of Emotion and Attention, 39–58.

Lavender, T., and Hommel, B. (2007). Affect and action: towards an event-coding account. *Cogn. Emot.* 21, 1270–1296. doi: 10.1080/02699930701438152

Marmolejo-Ramos, F., Arshamian, A., Tirado, C., Ospina, R., and Larsson, M. (2019). The allocation of valenced percepts onto 3D space. *Front. Psychol.* 10:352. doi: 10.3389/fpsyg.2019.00352

Marmolejo-Ramos, F., Cousineau, D., Benites, L., and Maehara, R.-o. (2015). On the efficacy of procedures to normalize Ex-Gaussian distributions. *Front. Psychol.* 5:1548. doi: 10.3389/fpsyg.2014.01548

Marmolejo-Ramos, F., Tirado, C., Arshamian, E., Vélez, J. I., and Arshamian, A. (2018). The allocation of valenced concepts onto 3D space. *Cogn. Emot.* 32, 709–718. doi: 10.1080/02699931.2017.1344121

Mather, M., and Knight, M. R. (2006). Angry faces get noticed quickly: threat detection is not impaired among older adults. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 61, P54–P57. doi: 10.1093/geronb/61.1.P54

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., and Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Pers. Soc. Psychol. Rev.* 9, 184–211. doi: 10.1207/s15327957pspr0903_1

Ossandon, J. P., Onat, S., and Konig, P. (2014). Spatial biases in viewing behavior. *J. Vis.* 14:20. doi: 10.1167/14.2.20

Parrott, W. G., and Sabini, J. (1990). Mood and memory under natural conditions: evidence for mood incongruent recall. *J. Pers. Soc. Psychol.* 59:321. doi: 10.1037/0022-3514.59.2.321

Phaf, R. H., Mohr, S. E., Rotteveel, M., and Wicherts, J. M. (2014). Approach, avoidance, and affect: a meta-analysis of approach-avoidance tendencies in manual reaction time tasks. *Front. Psychol.* 5:378. doi: 10.3389/fpsyg.2014.00378

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* 62, 1457–1506. doi: 10.1080/17470210902816461

Rinck, M., and Becker, E. S. (2007). Approach and avoidance in fear of spiders. *J. Behav. Ther. Exp. Psychiatry* 38, 105–120. doi: 10.1016/j.jbtep.2006.10.001

Rothkopf, C. A., Ballard, D. H., and Hayhoe, M. M. (2016). Task and context determine where you look. *J. Vis.* 7:16. doi: 10.1167/7.14.16

Rowe, G., Hirsh, J. B., and Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proc. Natl. Acad. Sci. U.S.A.* 104, 383–388. doi: 10.1073/pnas.0605198104

Schwager, S., and Rothermund, K. (2013). Counter-regulation triggered by emotions: positive/negative affective states elicit opposite valence biases in affective processing. *Cogn. Emot.* 27, 839–855. doi: 10.1080/02699931.2012.750599

Schwarz, N. (1990). "*Feelings as information: informational and motivational functions of affective states*," in Handbook of Motivation and Cognition: Foundations of Social Behavior, Vol. 2, eds E. T. Higgins and R. M. Sorrentino (New York city, NY: The Guilford Press), 527–561.

Shapiro, K. L., and Lim, A. (1989). The impact of anxiety on visual attention to central and peripheral events. *Behav. Res. Ther.* 27, 345–351. doi: 10.1016/0005-7967(89)90004-1

Shapiro, L. (2011). *Embodied Cognition*. Routledge. doi: 10.5840/philtopics201139117

Sharbanee, J. M., Hu, L., Stritzke, W. G. K., Wiers, R. W., Rinck, M., and MacLeod, C. (2014). The effect of approach/avoidance training on alcohol consumption is mediated by change in alcohol action tendency. *PLoS ONE* 9:e85855. doi: 10.1371/journal.pone.0085855

Slaby, J., Paskaleva, A., and Stephan, A. (2016). Enactive emotion and impaired agency in depression. *J. Conscious. Stud.* 20, 33–55. Availabe online at: https://www.ingentaconnect.com/content/imp/jcs/2013/00000020/f0020007/art00003

Stephan, A., Walter, S., and Wilutzky, W. (2014). Emotions beyond brain and body. *Philos. Psychol.* 27, 65–81. doi: 10.1080/09515089.2013.828376

Stoykov, M. E., Corcos, D. M., and Madhavan, S. (2017). Movement-based priming: clinical applications and neural mechanisms. *J. Motor Behav.* 49, 88–97. doi: 10.1080/00222895.2016.1250716

Tatler, B. W. (2007). The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *J. Vis.* 7:4. doi: 10.1167/7.14.4

Thomas, R. C., and Hasher, L. (2006). The influence of emotional valence on age differences in early processing and memory. *Psychol. Aging* 21, 821–825. doi: 10.1037/0882-7974.21.4.821

Tomaszczyk, J. C., Fernandes, M. A., and Macleod, C. M. (2008). Personal relevance modulates the positivity bias in recall of emotional pictures in older adults. *Psychon. Bull. Rev.* 15, 191–196. doi: 10.3758/PBR.15.1.191

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5

Veenstra, E. M., and de Jong, P. J. (2011). Reduced automatic motivational orientation towards food in restricting anorexia nervosa. *J. Abnorm. Psychol.* 120, 708–718. doi: 10.1037/a0023926

Wadlinger, H. A., and Isaacowitz, D. M. (2006). Positive mood broadens visual attention to positive stimuli. *Motiv. Emot.* 30, 87–99. doi: 10.1007/s11031-006-9021-1

Wadlinger, H. A., and Isaacowitz, D. M. (2011). Fixing our focus: training attention to regulate emotion. *Pers. Soc. Psychol. Rev.* 15, 75–102. doi: 10.1177/1088868310365565

Walter, S. (2014). Situated cognition: a field guide to some open conceptual and ontological issues. *Rev. Philos. Psychol.* 5, 241–263. doi: 10.1007/s13164-013-0167-y

Wiers, C. E., Kühn, S., Javadi, A. H., Korucuoglu, O., Wiers, R. W., Walter, H., et al. (2013). Automatic approach bias towards smoking cues is present in smokers but not in ex-smokers. *Psychopharmacology* 229, 187–197. doi: 10.1007/s00213-013-3098-5

Wiers, C. E., Stelzel, C., Park, S. Q., Gawron, C. K., Ludwig, V. U., Gutwinski, S., et al. (2014). Neural correlates of alcohol-approach bias in alcohol addiction: the spirit is willing but the flesh is weak for spirits. *Neuropsychopharmacology* 39, 688–697. doi: 10.1038/npp.2013.252

Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., and Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach

bias for alcohol and improves treatment outcome. *Psychol. Sci*. 22, 490–497. doi: 10.1177/0956797611400615

Wiers, R. W., Gladwin, T. E., Hofmann, W., Salemink, E., and Ridderinkhof, K. R. (2013a). Cognitive bias modification and cognitive control training in addiction and related psychopathology: mechanisms, clinical perspectives, and ways forward. *Clin. Psychol. Sci*. 1, 192–212. doi: 10.1177/2167702612466547

Wiers, R. W., Gladwin, T. E., and Rinck, M. (2013b). Should we train alcohol-dependent patients to avoid alcohol? *Front. Psychiatry* 4:33. doi: 10.3389/fpsyt.2013.00033

Wiers, R. W., Rinck, M., Dictus, M., and van den Wildenberg, E. (2009). Relatively strong automatic appetitive action-tendencies in male carriers of the OPRM1 G-allele. *Genes Brain Behav*. 8, 101–106. doi: 10.1111/j.1601-183X.2008.00454.x

Wilming, N., Harst, S., Schmidt, N., and König, P. (2013). Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLoS Comput. Biol*. 9:e1002871. doi: 10.1371/journal.pcbi.1002871

Witt, J. K., and Proffitt, D. R. (2008). Action-specific influences on distance perception: a role for motor simulation. *J. Exp. Psychol*. 34, 1479–1492. doi: 10.1037/a0010781

Xing, C., and Isaacowitz, D. M. (2006). Aiming at happiness: how motivation affects attention to and memory for emotional images. *Motiv. Emot*. 30, 243–250. doi: 10.1007/s11031-006-9032-y

Zhou, Y., Li, X., Zhang, M., Zhang, F., Zhu, C., and Shen, M. (2012). Behavioural approach tendencies to heroin-related stimuli in abstinent heroin abusers. *Psychopharmacology* 221, 171–176. doi: 10.1007/s00213-011-2557-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 23 April 2021
doi: 10.3389/fpsyg.2021.599939

Taking Situatedness Seriously. Embedding Affective Intentionality in Forms of Living

*Imke von Maur**

Department of Philosophy of Cognition, Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany

Situated approaches to affectivity overcome an outdated individualistic perspective on emotions by emphasizing the role embodiment and environment play in affective dynamics. Yet, accounts which provide the conceptual toolbox for analyses in the philosophy of emotions do not go far enough. Their focus falls (a) on the present situation, abstracting from the broader historico-cultural context, and (b) on adopting a largely functionalist approach by conceiving of emotions and the environment as resources to be regulated or scaffolds to be used. In this paper, I argue that we need to *take situatedness seriously*: We need (a) to acknowledge that emotions are not situated in undetermined "contexts" but in concrete socio-culturally specific practices referring to forms of living; and (b) to agree that not only are context and emotions *used* for the sake of something else but also that the meaning-disclosive dimension of affective intentionality is structured by situatedness as well. To do so, I offer a multidimensional approach to situatedness that integrates the biographical and cultural dimensions of contextualization within the analysis of situated affective dynamics. This approach suggests that humans affectively disclose meaning (together) which is at once product and producer of specific forms of living – and these are always already subjects of (politically relevant) critique.

Keywords: situatedness, affective intentionality, practice, form of living, habit, affective biography, socially extended mind

OPEN ACCESS

Edited by:
Leon De Bruin,
Radboud University, Netherlands

Reviewed by:
Gunnar Declerck,
University of Technology of Compiègne, France
Lucy Osler,
University of Copenhagen, Denmark

***Correspondence:**
Imke von Maur
imke.von.maur@uni-osnabrueck.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *28 August 2020*
Accepted: *06 April 2021*
Published: *23 April 2021*

Citation:
von Maur I (2021) Taking Situatedness Seriously. Embedding Affective Intentionality in Forms of Living. Front. Psychol. 12:599939. doi: 10.3389/fpsyg.2021.599939

INTRODUCTION

A political caricature might amuse one person, leave another unmoved, give rise to outrage in another, and prompt thoughts of murdering the caricaturist in a fourth. Some people feel pure anger while filling out a form offering a third box between "male" and "female," while others feel relief when ticking that box. The release of the newest Thermomix elicits great excitement in many, whereas others can only shake their heads about this way of "cooking," while a few might exist who cannot but be indifferent about this, because they do not even know what a Thermomix is. These cases are not abstract and sterile examples from and for textbooks. They are ways in which humans affectively disclose meaning and thereby do not only make up their own worlds, but the worlds of other humans as well. If trans persons are confronted with hate, disrespect and even the denial of their identities and rights; if, on a societal level, the practice of cooking gets lost because whole cultures following "food trends" lose the capacity of that craft; if a teacher gets beheaded because of discussing Muhammad caricatures in class, 5 years after journalists were murdered for

publishing one in Charlie Hebdo – the emotions involved in such kinds of world-making need to be understood and evaluated. But how can these emotions or the absence of such be explained and how can we assess which affective reaction is appropriate? The thesis of the present paper is that there are emotions which can neither be understood nor be normatively assessed without reference to what I call "forms of living"[1]. Setting out this thesis, I *take seriously situated approaches* to emotions in this paper and develop a *multidimensional approach to situatedness*.

The "situatedness paradigm" of affectivity can be seen as a similarly influential refocusing like the "cognitive turn" within both the psychology and the philosophy of emotions in the 1960s. The framework of situatedness, which has already been well established for cognitive processes (Robbins and Aydede, 2009 and Newen et al., 2018 for an overview), got transferred to the affective realm (Wilutzky et al., 2013; Stephan et al., 2014; Krueger and Szanto, 2016 or Stephan and Walter, 2020 for an overview): Emotions are no longer regarded as purely private affairs of an isolated subject, but as phenomena which are inevitably contextual. Instead of focusing on individual agents and unidirectional episodes of emotions (a particular emotion type being directed at a concrete object), situated accounts investigate affective phenomena which unfold *between* individuals and their social and material environments in dynamic processes.

The impetus of theories of situated affectivity – namely overcoming an individualistic or even intrapsychic paradigm of emotions – is of great import. Yet, while the emphasis on the significance of embodiment and environment for understanding affectivity is right and necessary, the pioneering situated approaches in the philosophy of emotions which provide the frameworks and conceptual toolbox for analyses do not go deep enough. They focus on the impact of body and environment on single affective episodes in a concrete moment while abstracting from the broader socio-culturally and historically specific biographical context (e.g., a jazz musician who regulates their emotions by means of their instrument or a marital quarrel in a given social setting; see Griffiths and Scarantino, 2009; Krueger, 2014; Stephan et al., 2014; Colombetti and Krueger, 2015). Additionally, these accounts mainly focus on the functional aspect of situatedness, viewing emotions as "to be regulated" and the environment as "to be used" (see Slaby, 2016 or Stephan and Walter, 2020, who call this the "user-resource-model"). What is missing is a conceptualization of the situatedness of *affective intentionality as disclosing meaning*: that humans represent their surroundings as being meaningful in a specific sense by means of their emotions. Or as Wittgenstein famously has it: "The world of the happy man is a different one from that of the unhappy man." Taking these two restrictions together, what is missing in the work on situated affectivity is to provide a conceptual framework for this affective way of disclosing meaning in its situatedness within socio-culturally specific practices. To be able to analyze this is of utmost importance for understanding and normatively assessing urgent and prominently discussed affective phenomena with political relevance like the ones mentioned above.

My multidimensional approach to situatedness integrates the concrete situation of affective dynamics within a broader context. Based on the assertion that it is not "context" (as an abstract variable) in which affective processes unfold but *concrete* socio-culturally and historically specific *practices* and *forms of living,* I argue that the specificity of such a practice and form of living systematically structures the characteristics and the content of emotions. To acknowledge this, we have to look beyond the concrete moment in terms of both time and space – we need to consider the affective biography of the feeling person as a product and producer of the specific ways in which body and environment affect the way in which emotions disclose realities. Without acknowledging this, we cannot adequately explain why certain affective processes unfold in the first place, how they are experienced, interpreted by the self and understood or even sanctioned by others and how to assess their appropriateness. The framework I develop aims at enabling an assessment of life-form specific structuring effects of *situated affective intentionality* – and, if necessary, at a politically relevant critique of situated affective intentionality. The aim of the paper is to open a new perspective for a politically engaged philosophy of affectivity. As such it provides an overview of, and wants to motivate, a new paradigm of situated affectivity. Achieving this aim requires that relevant aspects and analyses of single cases cannot be discussed in all details and depths – this paper rather is meant to offer a framework for such.

In the section "Affective Intentionality in Life-Form Specific Practices: 'Little Worlds,'" I introduce the concept of "little worlds" to denote the context in which affective intentionality is situated as structured by concrete practices which refer to forms of living. To denote the content humans disclose via affective intentionality, I introduce the term "meaningful Gestalt." This content is only intelligible against the background of the practices and forms of living which again make intelligible the "little worlds." In the "Situatedness I: Synchronic-Local Perspective" section, I adopt what I call a "local-synchronic" perspective on affective intentionality. This means looking at the present moment, at concrete affective dynamics between individuals and/or the material environment and the impact of such contextual factors on the characteristics and content of affecting and being affected. Importantly, I conceive of the context and the emotions not in functionalist terms but (a) in terms of meaning disclosure and (b) in their practice-specificity. In the "Situatedness II: Diachronic-Global Perspective" section, I adopt a "global-diachronic perspective," i.e., I focus on the intertemporal dimension of life-form specific embeddedness – the "affective biography" of an individual. Additionally – this is the "global" feature of that perspective – I consider socio-cultural factors which lie beyond concrete local, present moment affective dynamics, namely encompassing historical and societal structures such as emotional fashions, ideologies or regimes tacitly shaping the present moment dynamics. While in the first two sections the individual is situated within a context,

[1]I do not claim this holds for any emotional reaction. There are also instantiations of emotions for the explanation of which this account does not help. For instance, very basic forms of trigger responses, like being afraid in front of a dangerous animal or being disgusted by rotten food, might be explained without reference to forms or living and seem to be better explained with reference to biology.

in the section "Situatedness III: Forms of Living Within the Subject: Normative Assessment of 'Little Worlds'" I invert this perspective and situate the life-form specific context within the feeling individual. To adopt this perspective is a consequence of my conviction that it is not sufficient to put "naked" subjects into a context and *afterward* analyze the effects of such a contextualization on the characteristics and content of the involved feelings. Rather, what the multidimensional situatedness framework of this paper indicates is that life-form specific situatedness structures the space of possibilities for affecting and being affected as well as the content and characteristics of actual affective engagements in a much more fundamental way[2]. Importantly, the historico-social, biographical (diachronic) and inverted dimensions I develop are not optional "add-ons" which can *also* be considered when thinking about situatedness. Rather, they necessarily structure the synchronic local perspective at issue, in the approaches providing the conceptual toolbox for situated affectivity – this is what is meant by "taking situatedness seriously." This shift in perspective has also serious consequences for a normative assessment of emotions. What we ultimately evaluate when we deem concrete ways of affective disclosure to be (in)appropriate are the forms of living they enable, sustain or prevent.

AFFECTIVE INTENTIONALITY IN LIFE-FORM SPECIFIC PRACTICES: "LITTLE WORLDS"

Emotions, according to the core assumption of situated approaches to affectivity, are not private affairs but embedded in or even extended by the socio-material environment. This insight is of great import. Yet, the frameworks and concepts for situated approaches to affective phenomena do not go deep enough in addressing specific ways of affective reality construction with political relevance. This restriction can be revealed by considering the two main ways of addressing the relationship between the feeling person and environment offered so far: (1) to conceptualize emotions as strategies for manipulating the environment (cf. Griffiths and Scarantino, 2009; Wilutzky, 2015) and (2) to focus on emotion regulation through an active manipulation of the environment (*scaffolding* and *niche construction*; Krueger, 2014; Colombetti and Krueger, 2015). A paradigmatic example for the first way is a marital quarrel in which emotional expressions are used to test how the other one reacts – to get information about the context (Griffiths and Scarantino, 2009; Wilutzky, 2015). The second way concerns the active manipulation of one's emotions by *making use of the material environment*, for instance by listening to specific music or going to a certain place such as a church versus a sports event (Colombetti and Krueger, 2015). In both ways, emotions and environment are (i) considered regarding their *functional* aspect – emotions as strategies or a resource to be regulated, and the environment as a functional niche or scaffold. And (ii) their situatedness primarily concerns the present perspective of concrete affective encounters in a given environment.

But emotions and environments are not only *used* for the sake of something else (epistemic, pragmatic, or regulative purposes) but structure the very space of possibilities in which *meaning* is disclosed by self and others. This (shared) disclosure of meaning takes place in a concrete situation, yet the *specificity* of this situation and *how* the contextual factors shape the affectively disclosed meaning is only understandable against the background of specific practices and forms of living. In order to understand how humans – *as* beings engaging in socially shared practices and living specific ways of life – disclose meaning (together) affectively we need concepts which denote this practice-relatedness for both, the meaning disclosed and the situational context being producer and product of such affectively disclosed realities. I call the former "meaningful Gestalts" and the latter "little worlds" and introduce them now before I can establish the multidimensional situatedness framework in the sections afterward.

In the same way as I build upon the framework of situatedness, I take for granted the insights of the work being done on *affective intentionality*, namely that via emotions humans disclose something about themselves and the world (see Goldie, 2000; Roberts, 2003; Slaby, 2008 among others). But it is crucial to clarify how I understand emotions and their content in the following. The content disclosed via emotions, namely their presenting the self and the world as meaningful in specific ways (as opposed to being merely internal physiological arousals) is what I call *a meaningful Gestalt*. With this concept I reject the idea that emotional content is reducible to well definable evaluative properties like "the dangerous" or "the beautiful" – what is called the formal object of an emotion. Based on the insight that this alone does not specify the concrete content of emotions well enough, Bennett Helm (2001) introduces the helpful concept of *focus* to the debate of affective intentionality to denote the background concern which makes intelligible the formal object of an emotion in the first place. This brings out the reasons for why I am afraid of an angry looking crowd passing my bicycle on the street – namely the meaning it has to me and my desire for it to remain intact. The occurrence of a specific type of an emotion in a specific situation (here: fear) is only understandable with reference to a more encompassing pattern: what is disclosed via emotions is embedded in a net of concerns and meanings of the subjects going beyond the present moment. I would also feel relief accordingly if the crowd just passes without even noticing my bike. Robert Roberts (2003) adds to this picture the concept of emotions as *concern-based construals*. Similar to how we visually perceive Gestalts in pictures for instance (like the Wittgensteinian duck-rabbit), we at the same time receive certain input and construe its meaning. This is why I understand "disclosure" not as a merely receptive term but as performative as well. It is not only one single aspect but a whole meaningful Gestalt that is brought into existence when we feel in certain ways, not only for

[2]This is also reflected upon in the work of Matthew Ratcliffe, 2008 and a crucial facet of what he calls "existential feelings." These are the conditions of the possibility for concrete emotional episodes to occur in the first place and thus structure the very space of possibilities for affectivity (see also Slaby, 2008). As Ratcliffe highlights that affective meaning making needs to be considered in a temporally extended manner it would be worth further studies to examine the socio-cultural structuring of existential feelings as well. For a practice-specific account of pre-reflective affective intentionality that builds upon a combination of Merleau-Ponty's normative notion of "being *toward* the world" and Heidegger's emphasis on the affective nature of Dasein see von Maur, 2018, chapter 2.

us, the feeling person, but for our environment as well. When a teacher is ashamed because they made a mistake in a lecture they do not only privately experience the situation as shame-worthy. They also construe a "reality" being shared with the students. This reality or: "little world" – as I call it and introduce in a moment – provides the space of possibilities for other affective reactions following the teacher's shame from their side as well as from the students'. The reality – the context – is a different one than if the teacher would have reacted with laughter. And this Gestalt they, as individuals, are aware of by means of their lived bodily experience (for a detailed version of this account of affective intentionality see von Maur, 2018, chapter 2).

The notion of a "little world" refers to Lugones's (1987) introduction of "worlds" to denote multiple ways of being and the navigation with and between them from a phenomenological perspective[3]. For instance a person might inhabit the "world" of academia, the particular idiosyncratic world of their family, of being a woman in a male-dominated workplace or that of "being a Latina." These "worlds" are experienced differently and demand different kinds of (affective) comportments. In different "worlds" subjects are more or less "at ease," as she claims; in some worlds we are able to "sink in" (Ahmed, 2006), whereas others are burdensome or even not opened up. Importantly, humans can inhabit different "worlds" while being in the same space:

> *"Both you and I might be in the same room of the same building in the same city, but if you are a white United States-born citizen and I am a Latin American born in Nicaragua, we will probably have different takes on what we experience in this room, and we will have different takes on our experiences depending on the dominant norms and practices of the particular situation and how we relate to these practices given the contexts which dominate our particular interpretations" (Ortega, 2001, p. 11).*

I adopt the term "little world" to highlight this specific normativity structuring the disclosed Gestalt (with the decidedly political implications). A concrete situation in which individuals disclose meaningful Gestalts (together) is describable as such a "*little* world." These can but do not have to coincide with more prevalent, enduring and dominant descriptions of society, such as gender or class, but can also be more idiosyncratic as I will later discuss, for instance the "little world" people disclose because of posting anything about their life in "social" networks. The teacher example above shows that it might be only once that a particular "little world" is disclosed, whereas others are enduring practices and more stable forms of living – such as being a climate activist or a fan of a particular basketball team. A "little world" can be occupied by just one person, but mostly the affectively disclosed meaning and normative structure refers to something socially shared. I might disclose my low-carb-superfood oatmeal alone at home as fulfilling my need for a healthy life, but this is in its specificity only intelligible against the form of living perpetuated through media, advertisement – i.e., a meaningful Gestalt materialized in social practices. Thus, I consider the situation in which affective intentionality takes place as a (shared) "little world," that is: as a practice-specific reality (at a concrete time and place) referring to a *form of living*[4]. Accordingly, the environment an emotion takes place in not only provides the frame for sending or getting social signals, to gain information or to dampen or amplify emotions, but it essentially involves individuals in specifically meaningful realities of life. In any concrete affective dynamic, something involves and touches the subjects. These realities are not enacted by individuals alone but in shared processes with others and material factors which are always already meaningful – meaningful, that is, against the background of forms of living.

Forms of living concern the cultural and social reproduction of human life. As such they do not only express themselves in different beliefs, value orientations and attitudes, but also materialize themselves in fashion, architecture, the justice system and ways to organize families (Jaeggi, 2014, p. 21). Importantly, forms of living are not personal, private affairs: they are not individual options but "transpersonally shaped forms of expression with public relevance" (ibid., p. 22). For instance, to adhere to or refuse a gender specific behavioral order is a disposition unavailable to individuals alone insofar as it relies on socially constituted patterns of comportments and meanings. The behavior of an individual inevitably affects not only those adhering to or refusing these patterns, but it also shapes the space of possibilities of others (ibid.). A boy, according to Jaeggi, is not able to cultivate his preference for pink clothing innocently for very long without being confronted with the circumstance that – in some societies – his taste is coded as "girlish" (ibid., p. 22 fn. 7).

In order to understand what it means to address the situational context in which emotions take place *as a life-form specific* context, a praxeological perspective is of help: because any form of living finds expression in specific practices and in turn, any practice refers to a specific form of living. Practices can be understood as performances of skilled bodies which are neither reducible to mechanical movements, nor conducted in the mode of reflexively or consciously intended actions. Someone who masters a specific practice *embodies* the knowledge, the skill; it is inscribed into the lived body in a way that the life form specific comportment becomes "second nature" (Scheer, 2012, p. 202). Practices are, at one level, composed of such individual performances. Yet these take place in, and are only intelligible against, the more or less stable background of other performances. Emotions are thus situated in contexts in which

[3]She says a "world" might be the "dominant culture's description and construction of life, including a construction of the relationships of production, of gender, race, etc.," (1987, p. 10) of for instance an actual society. It must not be of a whole society though but can also be "a construction of a tiny portion of a particular society. It may be inhabited by just a few people" (ibid.).

[4]I use this term in connection to Ludwig Wittgenstein's (1953) "form of life." This concept and also his work on blind rule following importantly highlight the pre-reflective nature of norm guided behavior. Also, Martin Heidegger's (1927) differentiation between unarticulated general understanding (*Verstehen*) and explicitly grasping (*Auslegen*) emphasizes that comportment is related to norm-guided practices but that following such norms is not a matter of reflection and deliberate action – to grasp hammering you already have to understand the general practice of carpentry (cf. Rouse, 2007, p. 643). Wittgenstein and Heidegger count as precursors of what later has been called practice theory (cf. Schatzki et al., 2001). Especially with his hermeneutics of Dasein, Heidegger influenced many authors working on humans as practically engaging, understanding beings-in-the-world. This implies a critique on individualistic, rationalistic, or representationalist ideas of human behavior (cf. especially the work of Charles Taylor and Hubert Dreyfus).

humans skillfully perform practices which are in their specificity intelligible against the background of concrete forms of living. Taking situatedness seriously involves investigating the influence of this kind of contextualization on the way humans are situated affectively in what I call "little worlds" – namely (shared) spaces of *complex meaningful Gestalts*[5]. In the following section, I zoom in on concrete affective dynamics to explore the life-form specific structuring effects of situatedness on phenomenal character and the disclosed content of affective intentionality in (i) interpersonal and (ii) socio-material practices.

SITUATEDNESS I: SYNCHRONIC-LOCAL PERSPECTIVE

With an emphasis on the reciprocity, flexibility and openness of affective dynamics, situated approaches focus on the exchange of signals for the means of relationship configuration (Griffiths and Scarantino, 2009) or for epistemic purposes (Wilutzky, 2015). But the back and forth of affective interactions can also be addressed regarding the shared construction of "reality." The social psychologist Wetherell (2012), for instance, takes into view such normative sequences of situated affective dynamics by means of conversation analysis:

> *"The positions taken up are responsive to what has gone before, and are often loosely paired with each other. The affective pattern is in fact distributed across the relational field and each partner's part becomes meaningful only in relation to the whole affective dance [. . .] We create contexts for others as we act. Then, in reply, the other we have addressed orients to what is taking shape and remakes the context again" (Wetherell, 2012, p. 87).*

In affective dynamics, patterns develop for possible emotional reactions built upon those the dialogue partner offers, so that the "little world" and the Gestalt change likewise in a metamorphic process. This transformation dynamic is not only an interpretative framework of outside observers but is *experienced* by the involved subjects *through* their lived body. This can be described as a "sensual metamorphosis" – to use a term by sociologist Jack Katz. In his book *How Emotions Work* (Katz, 1999) Katz documents several studies he conducted about car drivers in a chapter called "Pissed off in L.A.". The fact that the reports analyzed are from Los Angeles is relevant. Driving a car in L.A. significantly differs from driving a car (as the general practice) in other contexts – for instance on a country road or on a highway in the Rocky Mountains. Also, anyone who has ever driven a car in Italy or France knows that driving and going postal – e.g., sounding one's horn – differs in frequency and intensity (i.e., in the affective involvement) a lot depending on the cultural setting. At first glance the scenes of outrageous car drivers seem to be characterized by the fact that they descend upon the person dramatically and unfold and progress in an uncontainable manner. This would support the common view according to which emotions are primarily (or even merely) an expression of internal physiological arousal of single individuals. But if driving the car is addressed as a bodily experienced socio-cultural practice it becomes visible that these affective processes do not develop like a chaotic hurricane but rather exhibit a specific *normative order*. Take this example from Katz:

> *"Lori, who is originally from Georgia but has lived in L.A. for many years, prefers public transportation but must drive here routinely. When 'a big new brown truck . . . decided to cut her off, Lori turns to the truck, 'What do you think you are doing? You know better than that!' She talks to herself and uses hand motions. She looks toward the driver in a sideways glance and then talks facing straight ahead . . . She does not want to lose her life over a driving dispute.' But after she goes through scolding motions 'she [can] drop it'" (Katz, 1999, p. 19; also quoted in Wetherell, 2012, p. 77).*

Because of the *established practice,* Gestalts are offered that are "worth freaking out over" – like tailgating, flashing headlamps, etc., which lead to typical emotional reactions expressed by screams of outrage, threatening gestures and mumbled (or loudly uttered) swear words.

This structure of affective dynamics cannot be explained by the established affective style between subjects who know each other well (as the so far established situationist approaches would do), but rather stems from the *shared practice* they are involved in, and the rules and norms which are known and accepted or refused (tacitly). Think of an escalating affective tumult emerging when those wanting to enter a train systematically block the doors and nobody can leave the train, or if a passenger realizes that someone else – supposedly wrongfully – is sitting in the seat they made a reservation for. Here affective dynamics emerge which – independently from the concretely involved individuals and their concerns – reveal an astonishingly intertemporal persistence in their patterns. The normative back and forth appears to be downright *scripted*[6]. There are roles for specific affective performances in which people slip in and out like professional actors.

Not only are other people part of affective processes, but also spaces, objects, infrastructures, etc., – the material "non-living" environment – build their context[7]. Freaking out while driving the car for instance co-depends on the way in which traffic is regulated:

> *"Those in cars whizzing toward us on the opposite side of the motorway or on the other side of the dual carriageway are rarely assholes. 'Assholeness' entirely depends on patterns of contiguity and common movement and, thus, occurs most often in relation to cars and drivers immediately in front of us and behind us heading in the same direction" (Wetherell, 2012, p. 88).*

[5]For a detailed account on "skillful coping" in this manner see especially the work of Hubert Dreyfus who also relates this to (background) practices (cf. Dreyfus and Wrathall, 2017). In his work as well as that of Charles Taylor (and their collaboration), also the epistemic picture influenced by Gestalt psychology (basic perception as being already meaningful) is a key issue (cf. Taylor, 2006). Both accounts as well as my approach developed here stand, in this regards, in theoretical debt to Martin Heidegger's hermeneutics of Dasein.

[6]For a detailed account on how emotions can be conceived of as following scripts see Eickers, 2019.

[7]See Malafouris, 2013 or van Dijk and Rietveld, 2017 for different approaches to explore the socio-material context and its impact on intentionality.

The car itself can even be interpreted as a physical extension of the self which enables specific ways of interaction with other vehicles (or the drivers). In this line, Katz even conceives of the car as being integrated in the body schema[8] of the driver. This could explain why primarily the drivers and not the co-drivers freak out and why driving an SUV feels different from driving a Smart. To consider cars and the contextual factors of being close or far away as material structuring factors on affective intentionality concerns how these factors impact the disclosure of another object (of the car driver, the whole situation, etc.). But also, the intentional objects are addressable from a practice-specific perspective. Here the concept of "affordances" is of help. James Gibson (1986) introduced the term for relational properties of objects which provide or prevent specific action-oriented offers – affordances – to the perceiver[9]. For instance, a chair is perceived as affording to be sat on or a piece of cake affords to be eaten. Making use of it for the realm of emotions, the concept of affordances concerns the phenomenological observation that some aspects in a situation have a specific "affective allure" (Rietveld, 2008, p. 977) or "affective power" (Romdenh-Romluc, 2013, p. 11) – they are felt as being salient in contrast to others and thus afford specific actions (see also Hufendiek, 2016 for a detailed approach to emotions as representing affordances). For the purpose of the present paper there is a relevant extension of Gibson's account, put forward by Allan Costall (2012) who suggests that we differentiate between ordinary and what he calls "canonical" affordances. The latter are distinctly concerned with the *socio-cultural background of practices* which make the affordance of an object intelligible:

> *"[S]uch affordances are situated not just in the 'current' behavior setting, but also in a more encompassing, shared and historically developed constellation – such affordances exist as they persist in shared and social practices [. . .] They exist as many individuals act on them in more or less appropriate ways, in the totality of practices that, together with other affordances, sustain them" (van Dijk and Rietveld, 2017, p. 3).*

In line with the key assumption of my multidimensional approach, the claim is that the relevant aspects of the environment of an individual in a concrete situation are only comprehensible insofar as they are considered as part of a more encompassing constellation of practices beyond the present moment (van Dijk and Rietveld, 2017). Material aspects are thus embedded in and comprehensible against the background of a conglomerate of practices too. The ordinary understanding of materiality as "pre-formed substances" (Orlikowski, 2007) has to be reconsidered accordingly and materiality and socio-cultural practice have to be seen as constitutively intertwined:

> *"[T]he social and the material are considered to be inextricably related – there is no social that is not also material, and no material that is not also social" (Orlikowski, 2007, p. 1437; also quoted in van Dijk and Rietveld, 2017, p. 4).*

The relationship between a practice and an affordance is according to van Dijk and Rietveld an example for such a relation of "constitutive entanglement":

> *"A specific practice and the affordance taking shape within it are interdependent and none of the two is prior [to] the other. Any affordance implies a practice which it realizes and any practice implies a landscape of available affordances" (2017, p. 4).*

To transfer this insight to the situatedness of affective intentionality as established so far, with a focus on the disclosure of practice-specific materiality, consider the following vignette:

Alex enjoys the first spring sun while shopping in Berlin. They are in the capital for an internship, but right now it's the weekend: leisure time. Alex already came across a variety of hip shops, bought trendy clothes and tested a fancy kale smoothie. While imagining, with a huge smile on their face, how to combine the new clothes and what to wear for the party tonight with their colleagues, Alex passes an impressive arrangement of gray blocks of stone. They feel the need to take a picture and share it on Instagram. A yoga pose, that would look great – Alex thinks. And in the next moment they ask a person to take a picture of them on the stone, one leg behind and the arm to the front. "Awesome!" Alex thinks happily, puts a hashtag below the picture and clicks "share." Filled with feelings of urbanity, creativity, inspiration, and freedom and a thrill of anticipation of the many likes and comments the picture will receive, Alex continues their shopping trip through Berlin.

How can Alex's emotions be explained without reference to the form of living their affective disclosure represents? Which relation holds between the properties of the stone blocks and Alex's reaction of happiness and enthusiasm? From an affordance perspective one could say that they perceive the stones as being "Instagram-able." Adding Helm's concept of *focus* we can specify that their happiness arises from the background concern which determines the meaningfulness of the object. But how can the background concern and the meaningfulness of the stones be described without reference to the socio-cultural practice of the very specific way of interacting on "social" media? Although it is true that these follow very specific normative rules which are permanently subject to subtle processes of change which are hard to understand for "outsiders" – there *is* something "at issue and at stake" (Rouse, 2002) that might escape being graspable by language, but that systematically structures the complex Gestalt that Alex discloses and the focus making the disclosed reality intelligible in the first place. This practice – referring to what I call the form of living of "posting" – structures the properties of a specific intentional object for different people as "post-able" (or Instagram-able, YouTube-able, Facebook-able, etc.), whereby the *concrete* Gestalts which are disclosed are possibly highly idiosyncratic[10]. A fashion blogger also presents the stone blocks

[8]The concept by Merleau-Ponty allows to see that gaining a new habit means to change one's body schema. For instance, a blind man's stick is integrated into the body schema: the blind man experiences the environment *via* the stick, they incorporate the stick and thus acquire the skill to inhabit the world in a different way than before (Merleau-Ponty, 1945, p. 176).

[9]"The *affordances* of the environment are what it *offers* the animal, what it *provides* or furnishes, either for good or ill. [. . .] I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment. [. . .] They are unique for that animal. They are not just abstract physical properties" (Gibson, 1986, p. 127).

[10]This hypothesis can be opposed from the very perspective from which I build it up. Especially *such* forms of living which are in a special way hip and fashionable, one could argue maliciously, lead to the perception of very *similar* Gestalts. From a

as "post-able" because they inhabit the practice of "posting," but a different Gestalt is disclosed – they see themself in a specific style, associated with possible advertisement partners, etc. A couple, in turn, wants to share their everyday life with "friends" on Facebook and takes a "partner-selfie" that should demonstrate (or even realize) happy moments and the narrative of the perfect relationship.

Importantly, this suggested *practice-specific affordance account* makes visible why certain objects, as opposed to others, even appear *as* objects for a certain affective disclosure – why *they* "pop-out" of a landscape of many possible affordances in that specific way (see von Maur, 2018, chapter 2.4 and chapter 4 for a detailed account on the pre-reflexive level of habitual affectivity)[11].

The "skillfulness" of affective intentionality, which is conceptualized in action- and goal-oriented functionalist terms in other situated approaches (cf. Wilutzky, 2015; Hufendiek, 2016), thus shifts on my perspective: The skillfulness dimension refers to the ability to affectively disclose what is "at issue and at stake" (Rouse, 2002) in a given practice. To reformulate Helm's concept of the focus from the perspective of the situatedness in life-form-specific contexts thus means to understand the concerns of the individuals against the background of what is "at issue and at stake" in a concrete situation relative to a specific practice and the norms constituting it. With this phrasing Joseph Rouse describes the normative element of practices, which is not reducible to either explicit rules or regularities, nor graspable or expressible through language.

> *"[W]hat a practice is, including what counts as an instance of the practice, is bound up with its significance, in terms of what is at issue and at stake in the practice, to whom or what it matters, and thus with how the practice is appropriately or perspicuously described" (Rouse, 2002, p. 175).*

> *"Our normative reach always exceeds our grasp, and hence what is at stake in practices outruns any present articulation of those stakes. [. . .] We are accountable to what is at stake in our belonging (causally and normatively) to the material-discursive world: our fate is bound up with what is at issue and at stake in our practices, although those stakes are not yet definitively settled – indeed, that is part of what it is for them to be 'at stake'" (Rouse, 2002, p. 25).*

In a practice-specific situation something is at issue *because* the interactants provide the context for the other one which is intelligible for the concrete other one or a relevant (in the sense of being familiar with the specific practice) community:

> *"[O]ne agent's situated environment and the possibilities it affords incorporate the activities of other agents as partially reconfiguring their shared surroundings. There is something at stake in intra-action with other agents, because its outcome shapes the intelligible possibilities for action and self-understanding by everyone involved" (Rouse, 2002, p. 21)*[12].

Someone who does not inhabit the practice of "posting" is not able to disclose similar Gestalts on pictures in forums or blogs affectively as someone who does. Someone not being fan of a "youtuber" (or even being unfamiliar with the existence of youtubers or the possibility of them being idols) is not able to disclose the Gestalt a fan discloses via being euphoric.

The interim result is that the context of a dynamically unfolding affective situation can be described as a specific "little world." The meaning which is disclosed in the form of complex Gestalts is co-constituted through the concerns of the involved feeling persons in relation to the practice. Life-form specific affordances affect us due to the incorporation of practice relevant norms and are thus always already meaningful and normatively structured with respect to practices and forms of living. Humans are "skilled" to disclose practice-specific normativity affectively and this skillfulness concerns the maintenance of the practices, the maintenance of specific "little worlds."

SITUATEDNESS II: DIACHRONIC-GLOBAL PERSPECTIVE

The synchronic situated perspective corrects an individualistic and decontextualized account of affective intentionality *spatially* by considering the concrete local environment of the feeling person. The diachronic perspective allows additionally to address an "*intertemporal*" dimension. Taking situatedness seriously now requires an integration of these perspectives in order to bring to light that, and explore how, affective intentionality in concrete encounters is structured not only by the people and artifacts present in that moment, but additionally by *the sedimented affective biography* which manifests in the practice and life form specific *emotion repertoire* a person acquires. The *emotion repertoire* is the set of meaningful Gestalts being available in a certain situation, given the learning history of the meanings of affectively relevant situations or cues (for a detailed account of emotion repertoires see von Maur, 2018, Chapter 4). Yet, taking situatedness seriously requires us going even further and considering a *global* dimension as well: Affective biographies differ depending on the era and culture in which they take place – namely the "cultural emotion repertoire."

critical perspective one has to consider the socio-culturally structured affectability and meaningful Gestalts in the mode of life of "das Man," as Heidegger calls it (see von Maur, 2018, chapter 4 for a detailed approach).

[11]I develop the concept "habitual affective intentionality" in order to explain in which way emotions are relevant for the epistemic goal of understanding within socio-culturally specific contexts. The concept allows to integrate the world-directedness, the situatedness, and the habitual dimension of affective phenomena. According to my account, the specificity of a concrete instantiation of affective intentionality is an irreducible way of world-disclosure, structured through socio-cultural embeddedness and through individual habitual "orders of feeling." Emotions, understood this way, are potentially defective for understanding processes because the habitual dimension can foreclose alternative ways of understanding and because it binds individuals to orders of feelings, allowing them to sustain their forms of life. Making these mechanisms visible allows us to think differently about potential solutions in order to overcome serious epistemic problems in everyday encounters.

[12]Rouse adopts this concept from Karen Barad (1996), who introduces it in order to avoid the implication of the term "interaction" that there are two definite and confined systems or individuals (cf. Rouse, 2002, ch. 8). For the same reason, Dewey and Bentley (1949) speak about "*trans*actional" rather than "*inter*actional," in order to avoid substantialist connotations of static entities (cf. Burkitt, 2014, p. 19). More recently, Shannon Sullivan (2001) takes up this notion in order to highlight the dynamic, co-constitutive relationship between organism and environment.

I firstly illustrate the diachronic dimension by taking emotional ontogenesis as one important sequence of the affective biography and by combining insights from social psychology (Parkinson et al., 2005) and philosophy (de Sousa, 1987). Already in the early stage of the affective biography, the ways of interacting with people and materiality structure life-form specifically how the world is affectively disclosed. From the beginning, the learning process of emotional meanings is a relational one: in face-to-face affective encounters, caregiver and child each react reciprocally to the gestures, facial expressions and vocalizations of the other. Through the specific feedback the child learns to ascribe meaning to the consequences of its behavior and ultimately to use it (which it initially unreflectively did) strategically. An illustrative example for this is called "coregulated behavior" (Parkinson et al., 2005, p. 237): the caregiver strongly holds the child in their arms such that it cannot move its own arms anymore. A successful coordination between the two would consist in the child trying to free its arms which causes the caregiver to lose their grip. If this does not happen, the child will experience frustration which can be interpreted as an early instance of anger. It learns to connect the whole situation of its frustrated need and the non-reacting caregiver with the resulting feelings which it will later identify and denote as anger. Such interaction contexts in which children learn to associate their reaction as an expression of particular emotions are what de Sousa (1987) calls "paradigm scenarios." In the context of a paradigm scenario, the instinctive reaction of a child to a stimulus becomes part of an emotion. Smiling or crying for instance will become an expression of joy or anger (de Sousa, 1987, pp. 285–286). The whole complex structure of emotions (intentional object, formal object, expression, etc.) is acquired, according to de Sousa, in a paradigm scenario. *Which* strategies and behavioral patterns a child acquires and uses continuously is dependent, according to Parkinson et al. (2005), on how the caregiver interprets the behavior of the child and how they react accordingly. A screaming newborn might be perceived by one person as being legitimate in its needs, whereas another person may interpret the same affective comportment as an expression of illegitimate stubbornness. Each will react differently to the child – and thus differently shape its emotion repertoire. In the first case it is likely that anger will be used as a means to have influence in interpersonal relations. In the second case it is more likely that anger will be recognized as a potential source for conflicts and thus only be expressed if the other one will not cooperate. The way in which the caregiver handles the perceived situation of the child is itself dependent on the resources which are available in the specific socio-cultural context of the person (Parkinson et al., 2005, p. 238). Even if the frustrated needs of the child are perceived as being legitimate, the necessary resources might be missing which would allow the fulfillment of its needs. Or the child is perceived as not being justified in their needs, but the caregiver does not see any other option to calm it down than by acting according to its will. Thus,

> *"[c]ulture affects the early consolidation of emotional responses at both an ideological and practical level. [.] Infants adapt to a preexisting social world, but do not simply soak up its influences like sponges. Instead, they negotiate ways of making practical or communicative use of whatever cultural resources are at hand" (Parkinson et al., 2005, p. 238).*

In a further developmental stage, emotions are not merely directed at the environment but can also have the relationship with a caregiver or object as an object. Typical phenomena of this stage of "secondary intersubjectivity" are joint attention and social referencing (ibid., p. 242). According to a study by Hornik et al. (1987), cited by Parkinson et al. (2005), 12 month old infants play less with a toy if the mother expressed disgust toward it before than in cases where the mother smiled or behaved neutrally. The infant thus seems to understand the caregiver's evaluation of other persons or objects. The meaning of such a situation – and thus the meaning of the emotion as well as its intentional object – is structured through the concerns of the child and the caregiver against the background of the shared practice, the "little world" that both enact together; and this practice-related relational aspect enters into the constitution of meaning of the emotion-object pairing getting a place in the emotion repertoire of the child.

In a community in which relevant linguistic conventions are shared, the growing child is eventually able to use symbols in order to influence others. Objects of emotions are thus no longer restricted to the present situation but can also be abstract or anticipated aspects. Such abstract meanings are highly dependent on the socio-cultural context. The enormous influence of the permanent confrontation of media-circulated advertisement on the development of the emotion repertoire of a child is especially remarkable here. Products acquire a place in a narrative – for instance in advertisement spots in the TV, in serials or movies, on posters, packages of sweets – which affect children in very specific ways. Following the theory of paradigm scenarios, the affective experience is connected to the meaning that this media representation delivered[13].

> *"Children do not even need to be directly exposed to this propaganda for the cultural message to filter through to them through social networks, shaping their desires, and satisfactions. Furthermore, the stickers, badges, costumes, and play-figures that are purchased for them convey messages about group membership that also carry emotional power" (Parkinson et al., 2005, p. 244).*

In practices, these emotional evaluations materialize themselves by the social environment dealing with the products in a specific way. Take friends in kindergarten or school who wear a certain kind of clothing, possess specific games, or know

[13]In this way of learning the meanings of emotions, the reciprocity highlighted by Parkinson et al. (2005) is distinctly restricted. The potential to affect that media exhibit does not only have a huge impact on children. Desires and affections are not only awakened (as if they have been present before and only need to be activated), but are rather brought into being in the first place. Often this has little to do with actual needs of the consumers. To escape this (affective) power is very hard to imagine in cases in which the individual has not developed a critical, distancing and reflexive stance toward consumism. For a detailed critique of media such as TV and their impact on (the affective repertoire of) children see for instance Bernard Stiegler's work on "taking care" (original "prendre soin") which highlights the need not to let alone children while consuming media and count on their alleged ability to resist. He pleas for a need to take care of them, meaning *inter alia* to teach a critical engagement with media (e.g., Stiegler, 2010).

the relevant music. These are as formative as the attitude of the parents with these things – prohibitions, consent, or critical utterances with respect to said objects shape the affectively disclosed meaningful Gestalt of the children. Again Parkinson et al. (2005) emphasize that children are no "cultural dupes" who blindly adopt anything their environment offers to them, but are able to use the available resources in accord with their concerns. Against the background of what I developed so far, this assertion seems to be too optimistic: the fact that possessing specific products is decisive for whether a child in kindergarten, school, or sports club belongs to the group or not is affectively effective to such a great degree that I can hardly imagine a child being able to defy. My formulated thesis above is that the "skillful dimension" of affective intentionality can be understood as the ability to disclose practice-relative "appropriate" normativity. Applied here, this would mean that a negative emotion with regards to life-form specific positively coded objects would be an explicit distancing from the norms relevant for a maintenance of this form of living. Yet, this *is* possible and even necessary in some cases, as I will illustrate in section "Situatedness III: Forms of Living Within the Subject: Normative Assessment of 'little worlds.'"

To bring together the developed pieces so far, consider the example of Alex once again. Alex is affectable by the stone blocks the way they are because of their affective biography and the resulting emotion repertoire. Conceive for instance another person, say Elli, who, contrary to Alex, is affected by the stone blocks with pure horror and sadness. This is due to *her* emotion repertoire: during her affective biography she, as the grandchild of a Holocaust survivor, has very sensibly been brought up with the relevant material and the respective meanings – in this case, the Holocaust memorial in Berlin, the meaning of which Alex does not know (accidentally). Alex must not have been in such a direct contact with the Holocaust herself in order to be affectable in the way Elli is. The claim here rather is that the different meaningful Gestalts being disclosed with respect to one and the same materiality cannot be understood properly by merely looking at the present moment. We need to take the diachronic dimension into account which is itself also a product of specific socio-culturally contingent circumstances. This "global" dimension of situatedness makes visible that also "*cultural* emotion repertoires" which differ between space and time need to be considered. For instance, my grandmother would not have been able to be affected by *anything* as being "Instagram-able," for the form of living of "posting" did not exist in the first place[14].

For the purpose of this paper, I will briefly demonstrate the operative efficacy of this dimension by considering how, for instance, different norms *about* emotions, belonging to cultural repertoires, shape the very act of affective disclosure. Importantly, cultural specificity does not (only) denote the difference between countries, nations, or continents but refers more encompassingly to shared systems of meaning that are anchored in specific socio-cultural milieus. "Culture" is understood accordingly as "learned systems of meaning, communicated by means of natural language and other symbol systems, having representational, directive, and affective functions, and capable of creating cultural entities and particular senses of reality" (D'Andrade, 1984, p. 116). Such norms for feelings direct the (affective) comportment of feeling subjects more implicitly than explicitly: internalized "cultural models" (Mesquita, 2007; Mesquita and Leu, 2007) guide the subject in identifying emotion-specific norms and demands in specific socio-cultural settings:

> *"Cultural models represent not just the normative, but more importantly the habitual; they lend meaning to our daily behavior. [. . .] The functionality of emotions within a socio-cultural context requires that they be coordinated with the specific cultural models" (Mesquita, 2007, p. 411).*

Such operationally effective cultural models especially manifest themselves in narratives *through which* one's own emotions, and those of others, are interpreted. This results from the specific way in which the person learned to talk and think *about* emotions – as a part of the relational process of affective biographies in which emotion meanings are learned through paradigm scenarios and then are picked up, changed and transformed throughout the course of life. For instance, the ideal of humans as self-determined rational individuals which are able to control their emotions in order to supposedly clearly, "cold-bloodedly," and factually make judgments and achieve knowledge is an example of a shared emotion culture (or even ideology) shaping the affective Gestalt disclosure of a given situation. This culture-specific narrative structures the interpretation of emotions only to the degree in which it has been acquired through the culturally situated affective biography. Think of the widespread assumption about the nature of emotions according to which there is a tension between their overwhelming power and the possibility of autonomous control. This assumption delivers a blueprint for interpreting one's feelings (retrospectively), for how they are spoken about and – this is the most interesting thesis – how they are experienced in the very moment of taking place. A subject then for instance interprets their outrage while driving the car – to come back to the example of section "Situatedness I: Synchronic-Local Perspective" – already in the moment it is happening, and not only retrospectively *through* the narrative which developed during her affective biography; namely, that the emotion overcomes them and that they actively need to control it to supposedly be "rational" again. Thus, the labels which a person can use in order to denote the experience of an emotion are not *prior* to the emotions and are *then* added to specific episodes of experience – like a post-it, as Sara Ahmed (2010a,b) formulates this insight. Rather, the labels shape the emotional experiences themselves[15].

[14]The idea that there are not only individual but also cultural emotion repertoires restricting the individual ones is for instance reflected in the concept of "emotional regimes" by William Reddy (2004), Barbara Rosenwein's (2002) concept of "emotional communities," or Raymond Williams'1977 "structures of feelings."

[15]See also Reddy, who, adapting the speech act theory of John L. Austin, talks about "emotives" (2004, p. 128): "A type of speech act different from both performative and constative utterances, which both describes (like constative utterances) and changes (like per-formatives) the world, because emotional expression has an exploratory and a self-altering effect on the activated thought material of emotion."

It is important to note here that the affective life of humans is not determined by one static emotion repertoire referring to one specific form of living. The diachronic dimension sketched here is meant to highlight the temporal plasticity of the affective dispositions of individuals. Emotion repertoires thus have to be conceived of as malleable and constantly changing. Also, a person even might have conflicting Gestalts at hand to be disclosed in the same moment – think of the tension experienced when you do not know yet whether you want to laugh or cry about someone telling you about a mistake you made. You might disclose the "little world" of being offended or the one of being thankful for the help. The notion of meaningful Gestalts and of "little worlds" entail the plasticity and the complex nature of ways of being in the world in specific "worlds." Humans can also "travel between worlds," as Lugones (1987) importantly discusses. As a politically relevant aim, she conceives of this as a needed capacity in order to understand the experience of others. Understanding other "little worlds" as well as questioning one's own, and dropping some in favor of others, are important capacities humans need to cultivate. This seems to be especially difficult, for the very mode in which these are operative is tacit and not explicitly reflected upon – as Al-Saji says with reference to Linda Martín Alcoff, "we see through our habits; we do not see them" (2014, p. 138).

We can see at this point that humans learn in socio-cultural feeling cultures to be affected and to affect others in specific ways, to ascribe meaning to these performances (by themselves and others), and to construct their current affective reality on the basis of this learning history. Thus, not only is a particular context always already normatively structured relative to socio-cultural practices, but the person themself is pre-figured in their specific affective "I can" (Al-Saji, 2014, p. 189). I will illuminate this perspective of a kind of "inverted situatedness," a consideration of the "environment within the subject" along with all its decidedly moral, political and societal implications in the final section.

SITUATEDNESS III: FORMS OF LIVING WITHIN THE SUBJECT: NORMATIVE ASSESSMENT OF "LITTLE WORLDS"

The subject situated in a context that is structured by a certain form of living and discloses a "little world" (with others) is a "product" of their affective biography: sedimented emotion repertoires restrict the space of possibilities for potential ways of being affected and affecting others. Yet, *individual* emotion repertoires are not only shaped by encompassing temporally and spatially specific *cultural* emotion repertoires but are even "socially extended" (Gallagher, 2013) or "invaded" (Slaby, 2016) by social structures: sociality is internalized and embodied in the subject's (affective) comportment. Forms of living do not shape the subject from the outside but are, in a sense, already *within* the subject. Exceeding the awareness and control of the subjects, forms of living thus make up their "little worlds" – sometimes even in ways conflicting with norms and values, and ultimately ways of being-in-the-world, that the subject would reflectively endorse. The concept of "situated affective intentionality" that I established in the previous sections allows us to deepen and illuminate the concept of shared "little worlds" from a decidedly normative perspective: the concrete realities being affectively brought into existence can now be made subject to normative assessment. The critique made possible here is at the same time potentially emancipatory in its epistemic dimension by making the subject aware of the tacit structuring of their world-disclosure. My "multidimensional situatedness framework" thus provides the ground to assess the appropriateness of emotions in a much deeper way than established accounts (i.e., fittingness, moral aptness or prudence; see Deonna and Teroni, 2012 or D'Arms and Jacobson, 2000 for an overview) – namely as one that is in the end evaluating different forms of living which specific emotions support or prevent.

In the context of theories of situated *cognition*, Shaun Gallagher (2013) claims that we need to adopt a political and critical perspective on phenomena within the research of situated cognition (and affectivity, as I will argue in the following). He suggests a "liberal interpretation" of the thesis of a socially extended mind, which goes beyond the classical examples of notebooks as potential extensions of memory functions. Gallagher claims that specific social practices (for instance, manipulating the decision-making process of people who should donate at charity events) structure cognitive processes, and that the mind is in this sense *socially extended*. The crucial point is, according to Gallagher, that we can easily imagine cases in which such a socio-normative structuring of mental processes is *not in the interest of those involved*. Against the background of this assumption, he pleads for a "critical twist" in existing research in cognitive science about the thesis of the social extension of the mind (ibid., also see Gallagher and Crisafi, 2009). This results in a wide-ranging change in perspective that I suggest adopting for situated affective intentionality. Such a change does not mean merely adding more or other factors as potential extensions of the mind, but rather the interest of investigation toward the epistemic object "affectivity" changes. Not only are the operative processes or questions about the location of emotions (in the head, in the body, in the environment) the subject of investigation, but rather, socio-material factors of lifeworld practice are to be considered in their structuring role (which is potentially subject to criticism).

One domain of practice for highlighting this perspective shift is the workplace. Criticizing the functionalist paradigm of situated accounts of cognition and affectivity, Jan Slaby (2016) analyses how the minds of white-collar workers are, as he calls it, "invaded" by culturally specific technical infrastructures or institutional practices. Slaby makes clear that the unquestioned idea in the paradigm of situated cognition and affectivity (that I called functionalist, and he denotes as the "user/resource model") runs the risk of overseeing structural effects which go beyond the personal grip as well as a one-sided positive utilization of environmental structures. To illustrate the perspective of an invasion of practice-specific affectivity into the individual, Slaby asks the reader to imagine themself to be an intern on their

In the same manner, Hochschild (1983) emphasized that not only the expression of emotions but also the experience of an emotion is shaped by societal convictions and norms about feeling(s).

first day of the job in a big company. The intern finds themself in an environment in which the colleagues talk to each other and behave in a way which is unfamiliar for the newcomer. This circumstance demands to learn more than the regular ways of working. In order to belong to the company, it is not sufficient to know how to do the job but to understand which ways of comportment in which manners and circumstances are appropriate and necessary – especially of the informal kind. To "become one of them" means first and foremost, Slaby argues, to get used to *affective comportments* and *affective styles* and to adopt them (ibid.). Such a process of habituation leads to the whole way of comportment becoming second nature such that the intern does not perceive them anymore *as* practice-specific demanded affective requirements and norms. What characterizes areas of "in-depth affective modulation" in general, such as corporate work spaces, is that they at the same time demand and lead to severe shaping effects on the personality, including affectivity, which "is profoundly framed and modulated so that the affective and emotional dispositions of an individual squarely fall in line with the interaction routines prevalent in these domains." (Slaby, 2016, p. 2) Crucial questions which are almost completely missing in the recent literature on situated affectivity[16] can be addressed and investigated in the context of my multidimensional approach. In which ways are such formative social domains operative, how do individuals become used to them, how do comportment and affective styles mix with these life-form specific processes? All these questions have a normative implication and open up much deeper reflection on the appropriateness of emotions than most accounts deal with. Taking up the example of the "little world" which is disclosed in an office illustrates the difference between the classical situated paradigm and my approach. Interactive technologies in this area (as environmental scaffolds) lead to the enlargement of working hours in areas which have been off times before (Slaby, 2016, p. 9) – "for instance, when office workers tend to be online and available for work-related communication night and day, no matter whether on weekends or during holidays" (ibid.). Individuals thus often do not actively decide in which way the environment modulates their affectivity, and these unconscious structuring effects invading from outside often even diametrically oppose the concerns of the feeling person. The "little worlds" which are established by life form specific affective intentionality are thus not neutral and equally preferable. There are worlds we should and worlds we should not disclose – dependent on the ways we aim to be in the world more generally. Highlighting the potentially negative impact of structures and practices on affectivity, the approaches of Gallagher and Slaby suggest that something from outside invades the subject, that something concrete intends to elicit specific processes within the individual. But driving cars, being a fan of a pop group, following food trends, or giving a talk at an academic conference are practices in which the specificity of the form of living structures the character and content of affective intentionality systematically, without being intended either from inside or from outside (as in the case of charity and seeking donation, in which the structuring of cognitive processes is intentionally aimed at). These structurings are *performed* – they become real by the fact that concrete individuals affect and are affected in a specific way. The discussed practices making up the form of living of "posting," the practices in office workplaces, as well as cultural standards about how to drive cars, already demonstrate this.

Even deeper though, our seemingly fundamentally personal desires are shaped by life-form specific practices. Thus, *which* "little worlds" we disclose is neither pure coincidence nor a solely private affair. Instead subjects *learn* for instance "what makes them happy" (Ahmed, 2010a,b) in their culturally specific and thus contingent affective biographies. According to the common picture we think that we are happy *because* that to which our happiness is directed is good. Contrary to this, Ahmed writes:

> *"[R]ather than say that what is good is what is apt to cause pleasure, we could say that what is apt to cause pleasure is already judged to be good. [. . .] Certain objects are attributed as the cause of happiness, which means they already circulate as social goods before we 'happen' upon them, which is why we might happen upon them in the first place" (2010a, 41).*

Thus, which "little worlds" appear to be attractive and thus are likely to be disclosed (together) affectively is fundamentally life-form specific: we "know" that champagne "tastes good," that wealth "makes us happy," and we associate our feelings with these objects according to this knowledge, according to the incorporated taste[17]. To drive a Porsche or SUV, to be "rich and famous," to possess the newest iPhone, or to wear the hippest fashion label, are in the same way already marked as objects of happiness *practice and life form specifically* – just as liking oysters, listening to the opera, or reading world literature are classified as "good taste."

For a normative assessment of emotions which takes situatedness seriously, the important implication is that not all emotions exhibit the value of "making happy," and thus the promise of happiness guides life in *certain* directions and not others. As an "emotional community" (Rosenwein, 2002) a family, like the work place, provides specific emotion repertoires, and refuses others. "Little worlds" are brought into existence, manifested, and transformed through affective dialogical practice. According to Ahmed, the family is not an object that is associated with happiness because it actually makes us happy but because the family is classified as a good, as an object to which positive affect sticks. To be loyal to the family goes hand in hand with the expectation of happiness. This orientation toward the object "family" influences the comportment extensively: "[Y]ou have to 'make' and 'keep' the family, which directs how you spend your time, energy, and resources." (ibid., p. 38). In a family, specific patterns of interaction and norms allow specific affections and prevent

[16]For a recent exception see Haq et al., 2020, who analyze radicalization processes through the lens of situated affectivity by making use of Slaby's concept of mind invasion.

[17]In the sense of Pierre Bourdieu (1979) who uses this to refer both to gustatory and aesthetic abilities being related to different habitus.

others. If we feel happy regarding these which are associated with happiness we are aligned: "we are facing the right way" (ibid., p. 37). But:

> *"We become alienated – out of line with an affective community – when we do not experience pleasure from proximity to objects that are already attributed as being good. [. . .] We become strangers, or affect aliens, in such moments. So when happy objects are passed around, it is not necessarily the feeling that passes. To share such objects (or have a share in such objects) would simply mean you would share an orientation toward those objects as being good" (Ahmed, 2010a, pp. 37–38).*

A subject who does not assimilate herself into the prescribed, learned construction of the meaning of feelings *already thereby* destroys the happiness of the others and is responsible for potential collapses of "little worlds." If a bad mood develops at the family table, for instance, the cause for this is seen to be the person who allegedly destroys the happiness of the family – the one who "kills the joy." By this, happiness is destroyed in several regards, not only because the situation not to be upheld in its "chastity," but also because the family is endangered in its status as a "happy object" – because the killjoy refuses their loyalty.

This line of thought now allows us to see that aligned ("fitting") emotions are necessary in order to sustain specific ways of interacting, thus: specific practices and forms of living. Not to feel aligned might make it impossible for some practices and forms of living to be upheld – it might end the existence of some "little worlds" and this has to be addressed normatively when it comes to the appropriateness of emotions. Emotions become important in the way that they allow or prohibit certain ways of living to be present – for good or bad. The way that meanings of emotions are learned, and how humans behave according to them, are structured through specific practices in a much more complex way as being visible if one abstracts from the multidimensional socio-structural situatedness that I have illuminated in this paper. Concrete emotions are explainable in their specificity because they allow the feeling person to partake in a specific form of living and to maintain it. Humans do not want to be "affect aliens" but rather strive for belonging, for fitting in. Against this background, the skillful dimension of affective intentionality concerns practice-specific responsivity allowing self and others to uphold the habitual "little worlds" and life-form specific realities.

CONCLUSION AND OUTLOOK

The present paper offers a framework that can address processes of (shared) meaning-disclosure in interpersonal and socio-material affective practices. In the course of their life, individuals negotiate meanings of emotions in relational affective processes with their socio-cultural environment. In accord with this, in a concrete situation a subject has *particular* Gestalts available for disclosing meaning. The meanings that objects acquire in this way are relative to forms of living and are in a crucial way at once *contingent* and *persistent*. They are *contingent* relative to the life-form specific paradigm scenarios in which an individual learns the meanings of emotions. The Gestalts an individual has at their disposal would be different if the person were raised in another epoch or culture, or if they had negotiated other meanings in relational processes with the relevant people. This means that the way in which humans are – or are not – affectable by particular affordances, and the Gestalts they can or cannot affectively disclose, are co-constituted by forms of living. These forms of living, in turn, are themselves the "products" of complex historico-cultural processes of becoming, and as such constantly subject to change. At the same time, the Gestalts an individual is or is not able to disclose exhibit a certain *persistence*: the way in which a person can be affected and affect others possesses some sort of perseverance and is often very hard to change. The way in which an individual construes reality becomes incorporated as second nature. Emotion-object pairings are dependent on the convictions about the "emotional value" of the objects which obtain in a given milieu. In this way the environment "invades" the repertoire of meaningful Gestalts – namely, how meaning is affectively construed. An emotion which seems inappropriate at first glance may actually manifest a resistance against emotional ideologies which ought to be called into question in the first place. If an investigation of affective intentionality only focuses on emotion types directed on particular objects with (un)fitting formal objects, then it abstracts from and is ignorant of *the reasons for the ascription* of these formal objects to the concrete things. As long as theorists operate with a repertoire of examples such as dogs and bears and their potential dangerousness, questions of the contingency and persistence of the meaning of "dangerousness" do not occur. But against the background of my multidimensional framework of situated affective intentionality, the assumption that this works the same way for examples like "what makes us happy" is inadequate. What *actually* makes us happy and what *should* do so is neither given by certain objects nor a question of personal preferences alone. It rather becomes comprehensible and criticizable against the background of the practices making intelligible the concerns which again explain the concrete emotions. A person is not enthusiastic about a thing like a Thermomix because of its supposedly objectively valuable properties nor because of their private preference of making any meal by heating and simultaneously mixing ingredients. They do so because they practice a specific way of "how one cooks" belonging to a certain form of living guided by a socially shared narrative – in contrast to people who do not cook at all or for whom cooking is a craft. In this paper I have dealt with emotions *in the life world practice*, with emotions beyond basic forms of trigger responses, with phenomena it makes sense to consider from a situated perspective. To take situatedness seriously means to explicate how life-form specific factors systematically structure the characteristics and content of affective phenomena and the "little worlds" thus brought into existence. Hence, concrete instantiations of affective phenomena are at the same time *producers* as well as *products* of socio-culturally specific practices and forms of living – and have to be normatively assessed as such. It is not only but especially vivid when looking at forms of living being guided by transphobic,

racist, sexist, or any other discriminatory emotion repertoire, that it matters *which* forms of living we sustain. The multidimensional framework developed here aims at contributing to and calling for a decidedly politically engaged situated approach to affective intentionality. It should provide the ground for a deeper analysis and a normative assessment of the effects of concrete practices and forms of living for our well-being, for what we deem to be lives worth living and for the political spaces we provide. It makes a huge difference *which* "little worlds" we disclose together affectively, and we need to direct attention to the severe and encompassing impact of this way of world-making.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I am grateful to the "critical emotion theory" network for a highly motivating discussion of an earlier draft of this manuscript: Ditte Marie Munch-Jurisic, Gen Eickers, Laura Luz Silva, Henrieke Kohpeiβ, Ruth Rebecca Tietjen, Marie Wuth, and Laurenzia Saenz. I also thank the Reading Club Affectivity of the University of Osnabrück, Rick Anthony Furtak for his careful reading and remarks, and the two reviewers for their constructive suggestions. My special thanks goes to Achim Stephan and Jan Slaby for their detailed feedback and their long-lasting support.

REFERENCES

Ahmed, S. (2006). *Queer Phenomenology*. Durham: Duke University Press.

Ahmed, S. (2010a). "Happy Objects," in *The Affect Theory Reader*, eds G. J. Seigworth and M. Gregg (Durham: Duke University Press), 29–51. doi: 10.1515/9780822393047-003

Ahmed, S. (2010b). *The Promise of Happiness*. Durham: Duke University Press.

Al-Saji, A. (2014). "A Phenomenology of Hesitation," in *Living Alterities. Phenomenology, Embodiment, and Race*, ed. Lee (Albany: State University of New York Press), 133–172.

Barad, K. (1996). "Meeting the Universe Halfway," in *Feminism, Science, and the Philosophy of Science*, eds L. Nelson and J. Nelson (Dordrecht: Kluwer), 161–194.

Bourdieu, P. (1979). *Distinction: A Social Critique of the Judgment of Taste*. New York: Routledge.

Burkitt, I. (2014). *Emotions in Social Relations*. London: Sage Publications.

Colombetti, G., and Krueger, J. (2015). Scaffoldings of the Affective Mind. *Philosop. Psychol.* 28, 1157–1176. doi: 10.1080/09515089.2014.976334

Costall, A. (2012). Canonical Affordances in Context. *Avant* 3, 86–93.

D'Andrade, R. (1984). "Cultural Meaning Systems," in *Culture Theory*, eds Shweder and LeVine (Cambridge: Cambridge University Press), 88–119.

D'Arms, J., and Jacobson, D. (2000). The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosop. Phenomenol. Res.* 61, 65–90. doi: 10.2307/2653403

de Sousa, R. (1987). *The Rationality of Emotions*. London: MIT Press.

Deonna, J., and Teroni, F. (2012). *The Emotions. A Philosophical Introduction*. New York: Routledge.

Dewey, J. and Bentley, A. (1949). *Knowing and the Known*. Boston: Beacon Press.

Dreyfus, H., and Wrathall, M. (2017). *Background Practices. Essays on the Understanding of Being*. Oxford: Oxford University Press.

Eickers, G. (2019). *Scripted Alignment: A Theory of Social Interaction*. Ph.D. thesis. Berlin: Freie Universität Berlin.

Gallagher, S. (2013). The Socially Extended Mind. *Cognit. Syst. Res.* 2, 4–12. doi: 10.1016/j.cogsys.2013.03.008

Gallagher, S., and Crisafi, A. (2009). Mental Institutions. *Topoi* 28, 45–51.

Gibson, J. (1986). *The Ecological Approach to Visual Perception*. New York: Psychology Press.

Goldie, P. (2000). *The Emotions. A Philosophical Exploration*. Oxford: Oxford University Press.

Griffiths, P., and Scarantino, A. (2009). Emotions in the Wild. *Robbins Aydede* 2009, 437–453. doi: 10.1017/cbo9780511816826.023

Haq, H., Saad, S., and Stephan, A. (2020). Radicalization through the Lens of Situated Affectivity. *Front. Psychol.* 11:205. doi: 10.3389/fpsyg.2020.00205

Heidegger, M. (1927). *Sein und Zeit [Being and Time]*. Oxford: Blackwell.

Helm, B. (2001). *Emotional Reason. Deliberation, Motivation, and the Nature of Value*. New York: Cambridge University Press.

Hochschild, A. R. (1983). *The Managed Heart. Commercialization of Human Feeling*. London: The University of California Press

Hornik, R., Risenhoover, N., and Gunnar, M. (1987). The effects of maternal positive, neutral, and negative affective communications on infant responses to new toys. *Child Dev.* 58, 937–944. doi: 10.2307/1130534

Hufendiek, R. (2016). *Embodied emotions. A Naturalist Approach to a Normative Phenomenon*. London: Routledge.

Jaeggi, R. (2014). *Kritik von Lebensformen*. Frankfurt: Suhrkamp.

Katz, J. (1999). *How Emotions Work*. London: University of Chicago Press.

Krueger, J. (2014). "Emotions and the Social Niche," in *Collective Emotions*, eds von Scheve and Salmela (New York: Oxford University Press), 156–171. doi: 10.1093/acprof:oso/9780199659180.003.0011

Krueger, J., and Szanto, T. (2016). Extended Emotions. *Philosop. Compass* 11, 863–878. doi: 10.1111/phc3.12390

Lugones, M. (1987). Playfulness, "World"-Travelling, and Loving Perception. *Hypatia* 2, 3–19. doi: 10.1111/j.1527-2001.1987.tb01062.x

Malafouris, L. (2013). *How Things Shape the Mind. A Theory of Material Engagement*. Cambridge: MIT Press.

Merleau-Ponty, M. (1945). *Phenomenology of Perception*. London: Routledge.

Mesquita, B. and Leu, J. (2007). "The Cultural Psychology of Emotions" in *Handbook of Cultural Psychology*, eds S. Kitayama and D. Cohen (New York: Guilford Publications), 734–759.

Mesquita, B. (2007). Emotions are Culturally Situated. *Social Science Information* 46, 410–415. doi: 10.1177/0539018407460030107

Newen, A., de Bruin, L., and Gallagher, S. (eds) (2018). *The Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press.

Orlikowski, W. (2007). Sociomaterial Practices: Exploring Technology at Work. *Organiz. Stud.* 28, 1435–1448. doi: 10.1177/0170840607081138

Ortega, M. (2001). "New Mestizas," "'World'-Travelers," and "Dasein": Phenomenology and the Multi-Voiced, Multi-Cultural Self. *Hypatia* 16, 1–29. doi: 10.1353/hyp.2001.0043

Parkinson, B., Fischer, A., and Manstead, A. (2005). *Emotions in Social Relations*. New York: Psychology Press.

Ratcliffe, M. (2008). *Feelings of Being*. Oxford: Oxford University Press.

Reddy, W. (2004). *The Navigation of Feeling. A Framework for the History of Emotions*. Cambridge: Cambridge University Press.

Rietveld, E. (2008). Situated Normativity: The Normative Aspect of Embodied Cognition in Unreflective Action. *Mind* 117, 973–1001. doi: 10.1093/mind/fzn050

Robbins, P., and Aydede, M. (eds) (2009). *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.

Roberts, R. (2003). *Emotions. An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.

Romdenh-Romluc, K. (2013). "Habit and Attention," in *The Phenomenology of Embodied Subjectivity*, eds R. Jensen and D. Moran (Cham: Springer), 3–19.

Rosenwein, B. (2002). Worrying about Emotions in History. *Am. Historic. Rev.* 107, 821–845. doi: 10.1086/532498

Rouse, J. (2002). *How Scientific Practices Matter. Reclaiming Philosophical Naturalism*. Chicago: The University of Chicago Press.

Rouse, J. (2007). "Practice Theory," in *Handbook of the Philosophy of Science. Philosophy of Anthropology and Sociology*, eds M. W. Risjord, P. Thagard, D. Gabbay, J. Woods, and S. P. Turner (Amsterdam: Elsevier), 639–681.

Schatzki, T., Knorr-Cetina, K., and von Savigny, E. (eds) (2001). *The Practice Turn in Contemporary Theory*. London: Routledge.

Scheer, M. (2012). Are Emotions a Kind of Practice (And is That What Makes Them Have a History)? A Bourdieuan Approach to Understanding Emotion. *Hist. Theory* 51, 193–220. doi: 10.1111/j.1468-2303.2012.00621.x

Slaby, J. (2008). *Gefühl und Weltbezug. Die menschliche Affektivität im Kontext einer neo-existenzialistischen Konzeption von Personalität*. Paderborn: Mentis.

Slaby, J. (2016). Mind Invasion. Situated Affectivity and the Corporate Life Hack. *Front. Psychol.* 7:266. doi: 10.3389/fpsyg.2016.00266

Stephan, A., and Walter, S. (2020). "Situated Affectivity," in *The Routledge Handbook of Phenomenology of Emotions*, eds Szanto and Landweer (London: Routledge).

Stephan, A., Walter, S., and Wilutzky, W. (2014). Emotions beyond brain and body. *Philosop. Psychol.* 27, 65–81. doi: 10.1080/09515089.2013.828376

Stiegler, B. (2010). *Taking Care of Youth and the Generations*. California, CA: Stanford University Press.

Sullivan, S. (2001). *Living Across and Through Skins. Transactional Bodies, Pragmatism, and Feminism*. Bloomington: Indiana University Press.

Taylor, C. (2006). "Merleau-Ponty and the Epistemological Picture," in *The Cambridge Companion to Merleau-Ponty*, eds T. Carman and M. Hansen (Cambridge: Cambridge University Press), 26–49. doi: 10.1017/ccol0521809894.002

van Dijk, L., and Rietveld, E. (2017). Foregrounding Sociomaterial Practice in Our Understanding of Affordances: The Skilled Intentionality Framework. *Front. Psychol.* 7:1969. doi: 10.3389/fpsyg.2016.01969

von Maur, I. (2018). *Die epistemische Relevanz des Fühlens – habitualisierte affektive Intentionalität im Verstehensprozess*. Osnabrück: Universität Osnabrück.

Wetherell, M. (2012). *Affect and Emotion. A New Social Science Understanding*. London: Sage.

Williams, R. (1977). *Marxism and Literature*. Oxford: Oxford University Press.

Wilutzky, W. (2015). Emotions as Pragmatic and Epistemic Actions. *Front. Psychol.* 6:1593. doi: 10.3389/fpsyg.2015.01593

Wilutzky, W., Stephan, A., and Walter, S. (2013). "Situierte Affektivität," in *Handbuch Kognitionswissenschaft*, eds Stephan and Walter (Stuttgart: Metzler), 552–560.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 25 May 2021
doi: 10.3389/fpsyg.2021.635993

Enacting Media. An Embodied Account of Enculturation Between Neuromediality and New Cognitive Media Theory

*Joerg Fingerhut**

Berlin School of Mind and Brain, Department of Philosophy, Humboldt-Universität zu Berlin, Berlin, Germany

OPEN ACCESS

Edited by:
Albert Newen,
Ruhr University Bochum, Germany

Reviewed by:
Regina E. Fabry,
Ruhr University Bochum, Germany
Marc Slors,
Radboud University Nijmegen, Netherlands

***Correspondence:**
Joerg Fingerhut
joerg.fingerhut@hu-berlin.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *30 November 2020*
Accepted: *15 March 2021*
Published: *25 May 2021*

Citation:
Fingerhut J (2021) Enacting Media. An Embodied Account of Enculturation Between Neuromediality and New Cognitive Media Theory. Front. Psychol. 12:635993. doi: 10.3389/fpsyg.2021.635993

This paper argues that the still-emerging paradigm of situated cognition requires a more systematic perspective on media to capture the enculturation of the human mind. By virtue of being media, cultural artifacts present central experiential models of the world for our embodied minds to latch onto. The paper identifies references to external media within *embodied, extended, enactive*, and *predictive* approaches to cognition, which remain underdeveloped in terms of the profound impact that media have on our mind. To grasp this impact, I propose an enactive account of media that is based on expansive habits as media-structured, embodied ways of bringing forth meaning and new domains of values. We apply such habits, for instance, when seeing a picture or perceiving a movie. They become established through a process of reciprocal adaptation between media artifacts and organisms and define the range of viable actions within such a media ecology. Within an artifactual habit, we then become attuned to a specific media work (e.g., a TV series, a picture, a text, or even a city) that engages us. Both the plurality of habits and the dynamical adjustments within a habit require a more flexible neural architecture than is addressed by classical cognitive neuroscience. To detail how neural and media processes interlock, I will introduce the concept of *neuromediality* and discuss radical predictive processing accounts that could contribute to the externalization of the mind by treating media themselves as generative models of the world. After a short primer on general media theory, I discuss media examples in three domains: pictures and moving images; digital media; architecture and the built environment. This discussion demonstrates the need for a *new cognitive media theory* based on enactive artifactual habits—one that will help us gain perspective on the continuous re-mediation of our mind.

Keywords: 4E cognition, architecture, artifactual habits, digital media, film, neuromediality, picture perception, predictive processing

INTRODUCTION

Media are the core currency of culture. Alongside images, texts, and sounds, new varieties of media (especially in digital form) profoundly shape human "pattern practices" (Roepstorff et al., 2010) across cultural domains. In these contexts, situated cognition is well-placed to examine how sociocultural niches scaffold and structure the mind. Yet paradoxically, media phenomena do not occupy a central place within the discourse on situated cognition. In this paper, I propose an understanding of our engagement with media artifacts based on a theory of habits. To explain such habits, I take my cue from enactivism and recent theories of embodied or radical predictive processing (Clark, 2013, 2015a). I demonstrate how such an understanding is needed to capture the disparate ways media artifacts engage us with their experiential models of the world. I center artifacts as an object of study due to their status as the most quintessential and enduring manifestations of human culture. Exploring a systematic media perspective about such artifacts ought to inform (and form an integral part of) situated cognition accounts of enculturation.

Enculturation is commonly understood as the acquisition of cognitive practices within sociocultural niches, covering ontogenetic levels of dynamic change that unfold across a lifespan. Such ontogenetic niches have been the focus in cognitive science and will be the focus of the present paper as well, as it mostly deals with media in what has been labeled "developmental" or "cognitive niches" (Stotz, 2010; Bertolotti and Magnani, 2017). Given its discussion of cultural evolution and cultural development, theoretical discourse on enculturation constitutes a significant addition to *embodied, embedded, extended,* and *enactive* (4E) cognitive science (Hutchins, 2011). Accounts for enculturation claim that "culturally mediated worlds in which we grow up and live are integral to how our brains achieve their functional capability" (Kirmayer et al., 2020, p. 6). This occurs holistically. For example, "cognitive integration" theories link the acquisition and entrainment of capacities for calculation to wider practices that encompass epistemic tools and representational systems within a culture (Menary, 2007, 2018).

While such accounts are theoretically invaluable, they typically focus on higher-level cognitive capacities (at least when considering paradigm cases). This includes capacities that are only made possible through cultural practices (Hutchins, 2008), as well as specific epistemic operations derived from certain tools and media. Among these are those relating to the capacities of reading, writing, memory, and mathematical cognition (Heyes, 2012; Menary, 2015; Fabry, 2018). Such accounts are not immediately concerned with broader questions regarding cultural tools and media, such as how they might afford novel, experiential models of the world. Moreover, the field does not sufficiently engage with human artifacts and media beyond notational systems and language. Other cultural artifacts, such as images and films, new and digital media, and the built environment, could be considered as equally central and pervasive insofar as they substantively structure our cognitive lives—they even permeate our perception and affectivity. By focusing on how we enact such artifacts, this paper aims beyond a single cognitive practice (made possible by the processes of enculturation) to explore how experiential domains are generated through embodied media habits.

Although the cognitive sciences routinely consult the theoretical traditions of philosophy and psychology, they often overlook relevant theoretical work in fields such as image science and media studies. This is unfortunate because media studies, especially, could be an important humanities companion to 4E cognitive science. As a field, media studies elucidates the inner operations and logics of different media systems. In doing so, it reveals the relevance of media's technological dimensions to our lives. After all, media are artifacts that expand our cognitive and experiential reach beyond traditional conceptions of the human senses. Media record, process, and transmit information. As media studies have shown, these basic operations developed over history in different cultural-technological niches and became implemented in specific forms. The prominent field of *media archeology*, for example, traces the trajectories of technological devices such as the typewriter, film, and computers (Kittler, 1999). Media theories therefore emphasize the material and technological underpinnings of media (Gane, 2005) while also showing how media amount to more than that. As Kittler asserts, "media determine our situation" (Kittler, 1999, p. xxxix). The guiding premise for the present paper, then, is that both 4E cognition and media studies capture the ways in which cultural artifacts shape our lives and minds.

I argue that the embodied habits and skills employed when engaging cultural artifacts constitute a central level of description (Fingerhut, 2020a). Habits are ways of acting. As such, they structure our perceptions, emotions, and thoughts. Habits are also expansive in three aspects: time, space, and the sphere of activity they afford. (a) Habits assemble tacit expectations within certain ecologies and therefore structure our future actions therein. Since those expectations have been shaped over time, habits link our current engagements also with our history of environmental coupling. In other words, they are *temporally expansive*. (b) Habits are co-constituted by our socio-cultural-technical environment, making them *locationally expansive* in the sense that, for example, media artifacts critically determine the way an engagement unfolds within a habit. (c) Interestingly, habits (which are often seen as exhibiting an inherent inertia) further exhibit a tendency to transcend themselves. They do so by adapting to novel circumstances or by unlocking new domains of interaction. This means they are *transformatively expansive*.

More specifically, this paper will explore the sensorimotor and body-schematic processes underlying artifactual habits with respect to pictures and cinematic productions—along with new (social and digital) media as well as the built environment (the lasting impact of which is re-mediated in our smart cities). Clearly, the processes constituting a habit are more complex and varied than such a focus can reveal. Higher cognitive processes also play a central role in the unfolding of skillful engagement, as is addressed by so-called 'vertical elements' in *meshed architecture*

accounts of skills (Christensen et al., 2016). Such processes go beyond the scope of the present paper, for its focus is not so much cognitive control but rather to what extent control over experiential engagement—in a more bottom-up fashion—is exerted by the medium and the interaction itself (Gallagher and Varga, 2020). The general emphasis is on medium-specific habits (i.e., how media habits differ from one another and how they are adaptive in specific media ecologies), the ways pervasive artifacts permeate our cognitive engagement down to the level of perception and affect (think of the impact of architecture, film, digital media), and how this pans out in encounters with specific media works (i.e., how we attune to media and how they entrain us in the *here and now*). With this emphasis, we can identify central elements of our engagement with the experiential models presented to us by media.

The first section of this paper briefly surveys some 4E cognition accounts that reference media to provide an understanding of the nature of the mental states that emerge when media engage human organisms. One focus will be the hybrid realization claims of the e*xtended mind*. Another focus will be the *enactive* nature of our mental states and the evaluative domains such an enactivism entails. I will subscribe to an enactive account of habits that highlights the active role of the body in bringing forth experiences with the purpose to extend this idea to media ecologies. Yet I will also discuss how this account can retain a focus on the hybrid material nature of the brain-body-nexus underlying such engagements and its structure (in broadly functionalist terms), which some variants of enactivism might reject.

The second section addresses the role of the brain in our media engagements more directly. Understanding the way our neuronal processes dovetail with media on different levels of the hierarchical processing of the brain (along with how this relates to the way information is recorded, processed, and transmitted in different media) could be taken up by *radical predictive processing* theories (Clark, 2013, 2015a). These theories give an account of the role that so-called designer environments (and media, in my understanding) play in the dynamics between the brain, body, and world. My focus here will be on active inference and design-guided bodily engagement. Within a habit, then, we can identify *neuromedial* elements that complement the unfolding of a media engagement. Such a framework presents itself as a theory of media as central experiential models of the world that need not be mirrored in the brain, but rather engage the brain-body nexus.

Section three recounts this idea and relates the situated mind to a general media theory. All this has implications for how we should conceive of our more specific media engagements. Section four therefore discusses examples of media engagement types (exploring also the mental states realized *within* an artifactual habit) and prepares the grounds for a *new cognitive media theory* based on what has been discussed before. It then associates these types to the medium-specific body schema we employ when engaging with film, to the capacity of *seeing-in* with respect to pictorial artifacts, to the ways new and digital media actively engage and predict their users, and – last but not least – to the understanding of the built environment as a media environment.

SITUATING MEDIA IN THEORIES OF THE MIND

Philosophy of mind is media theory. This is true in a general and rather trivial sense. What reaches our mind is mediated by our body-brain nexus and habits of interacting with the environment that we have acquired over time. Within a relational, situated philosophy of mind, the central function of the brain is one of a "mediating organ" (Fuchs, 2011). This organ facilitates engagements between an agent and the world, with those engagements themselves now gaining center stage for an understanding of the mind. But what seems trivially true does not translate easily into a theory. This is because a media perspective could erroneously suggest that the mind is a receiver that exists outside of the mediating apparatus—a position I argue against. In the following paragraphs, I will not explore the general concept of mediation, though, but rather focus on the role that external media play in situated cognition accounts and that might shed a light on the relational nature of our mind.[1]

A Mixed-Media, Deterritorialized Cognitive Science

External media have been most prominently referenced in theories of the extended mind (EM). These theories argue that media artifacts, whether a handwritten notebook or an iPhone, could be taken as literal parts of the machinery that realizes mental states (under specific circumstances, such as the reliability, trustworthiness, and accessibility of the external device). Clark and Chalmers' (1998) perennial thought example describes a notebook taking over the memory function of the brain of an Alzheimer's patient named Otto, substituting what would otherwise be carried out by neural realizers in healthy individuals. Beliefs therefore supervene upon a hybrid brain-artifact structure in Otto.[2]

In other writings, Clark emphasizes that external structures may not gain a central role in co-constituting cognition if they did not significantly complement what the brain-body nexus can do on its own: "external structures function so as to *complement our individual cognitive profiles* and to diffuse human reason across wider and wider social and physical networks whose collective computations exhibit their own special dynamics and properties [emphasis added]" (Clark, 1997, p. 179, 1998). Sutton (2010), who refers to such accounts as "second wave" EM, spearheads exograms (Donald, 1991) and the idea of exosomatic memory to drive home the point of complementarity. This latter notion

[1] As such, media concepts have already helped to structure some central debates in analytical philosophy of mind and consciousness. These include discussion of the analog or digital content of mental states, Dennett's rejection of any identifiable or special neural medium of consciousness, and his claim that consciousness is "fame in the brain" unbound from any specific medium (Dennett, 1993, 2001). Clearly, media theorizing can be fruitful for a heuristic of the mind. However, this paper is more immediately concerned with external media and the role they might play in constituting mental states.

[2] Much has been said about the extent to which inner neural and outer media processes must have similar processing properties to warrant parity of treatment. For a recent take on this, see Wheeler (2019). For a critical view on whether this introduces a mark of the mental based on properties of inner processing, see Di Paolo (2009).

is also core to aforementioned "cognitive integration" theories.[3] Exograms are external media storage devices, such as written books, images, libraries, databases. These devices do not simply mimic neuronal memory processes (engrams). Instead, they exhibit properties that inner processes typically lack: reliance, transmittability, reorganization, and so on. It is therefore the combination of inner and outer formats that was beneficial in such cases and which enables the human mind, as compared to a species exhibiting a more limited range of such combinations, to achieve novel and exciting things.

As the hybrid structures of EM suggest, mixing media might generally be advantageous. Clark (2019) refers to *DeepMind* or Differentiable Neural Computers, which are highly evolved machine learning systems. These successfully perform tasks by employing a so-called read-write unit that enables them to externalize certain processes in a different media format (by writing them out), thus giving them sensorimotor access to stable yet modifiable external storage elements (Clark, 2019, p. 272). This describes a cognitive solution that uses engrams and exograms alike. Such an artificial system might seem a rather alien example (albeit one that gains significance when we think of the effects of AI, ubiquitous computing, and the digitalization of our life world). Yet the example demonstrates how mixed systems, understood as one media system exploiting another, jointly constitute better cognitive solutions.

Sutton suggests also a third wave of EM: for human brain-body-artifact interaction we could consider dynamic "shifting networks of heterogeneous components temporarily clustered or clumped together in contingent coalescence" (Sutton, 2010, p. 194). This has further consequence for how we should study cognition:

> If there is to be a distinct third wave of EM, it might be a *deterritorialized cognitive science* which deals with the propagation of deformed and reformatted representations, and which dissolves individuals into peculiar loci of coordination and coalescence among multiple structured media [emphasis added]. (Sutton, 2010, p. 213).

With such a wave, we would study series of transformations occurring in interactions between human organisms and artifacts as temporal integrations. These integrations allow for de- and reformations as part of the cognitive process, before then fading out again.

Given the perspective on media proposed here (as a description of how we attune to media), it is tempting to follow a third wave of EM that highlights fleeting "soft" or "transient assemblies" (Clark, 1997, pp. 42–45, 2016, p. 150). This is because the cognizing organism need not be the center of control nor the sole focus when it comes to the kind of information processing involved. The organism also cannot claim agency in such assemblies (Kirchhoff, 2012). Consider smartphones and the other touchscreen devices that we carry around with us: they entrain us when we watch a video, for example, by providing a filmic exploration within a respective media-specific succession of frames (aided by sound and music). Yet they can also re-direct us to their surfaces, such as when we receive a message. Whether in entrainment or in the switching of attention, the activity is elicited and structured by the multi-media device. Similarly, consider how so-called smart cities of today aim to engage us (often also via screen-based media): they steer our movements and elicit cognitive processes by nudging our behavior and using their own algorithms to engage us (they do so more actively than traditional architecture, which already engages us by guiding our embodied exploration of space).

These examples make obvious that something else might be required beyond transient assemblies. In order to capture more fully the nature of our media engagements, we need to identify the central constraints that determine a specific kind of media engagement. We therefore have to focus on recurring media or artifact coalescences and the ensuing structured interactions they elicit. While it is true that an encompassing theory of enculturation also has to understand what it means for real-time coalitions of organism and artifact to mix and dovetail, it is as central to relate such media-mixing and reformatting of information to the more enduring habits sustained in specific media ecologies. Those habits determine our engagement with pictures, screen-based media, and the built environment, etc. Rather than therefore fully *deterritorializing* cognition (as third wave EM seems to suggest), this rather requires an enhanced focus on the cultural contexts that provide structure along with the recurring ways pervasive artifacts entrain us. What therefore is required is the mapping of multiple (often dormant) skills and habits that are constantly re-activated and re-negotiated upon exposure to a media environment, which I will address below.[4]

None of the above waves of EM amount to a theory of cognition on their own. Instead, they present some arguments that should inoculate us against simply assuming that cognitive processes are confined to the skull and skin and exclusively realized in a specific neural medium. The hybrid realization view of EM has relevance for the present paper as an epistemic claim: by giving up the focus on the locally instantiated brain-body, cognitive science should be rewarded with extra explanatory power and parsimony. We can track how in certain media engagements, organism and artifact jointly explore a content. The brain does not have to mirror the operations of the medium, but simply to latch onto them (as I will explore in section "The Radically Predictive Brain"). With respect to mental states such as beliefs, EM attributes a co-constitutive role to external media. But the hybrid realization view that underlies this move also connects with a broad functionalist commitment in EM (Wheeler, 2012). This commitment is not shared by other Es, as we will see shortly with respect to enactivism. They claim that what is central to

[3] See Menary (2007). For the human organism, the benefits of cognitive integration is that it enables us to do things "we otherwise could not do and [in] the transformation of existing abilities, making us smarter and better at difficult and demanding cognitive tasks" (Menary, 2018, p. 197).

[4] The dynamic exploitations of different media across brain-body-culture boundaries then unfold *within* such habitual engagements. This paper takes up the differences *between* media-habits as well as the rules of engagement *within* a habit (i.e., the specific unfolding of cognitive processes within a skill). I will not focus on what could be considered our meta-habits of (wittingly or unwittingly) choosing different media resources for engagement. Still, the latter has become a central focus in understanding media economies that compete for our attentional resources (Crogan and Kinsley, 2012).

cognition is not captured well by a computational description of information-processing (and thoroughly misrepresented when relying on representations).

Another well-known challenge to EM comes from internalists such as Adams and Aizawa (2001). This challenge grants that extra-bodily elements may indeed *cause* cognitive processes, but that we cannot infer constitution from causation. For our media cases, the question is: do operations outside the organism *co-constitute* the exploration and bringing forth of world models (such as sensorimotor loops structured by the filmic medium or those co-processed by artificial computations in virtual reality setting)? Or is it that the brain only *causally depends* on such media-body-brain couplings and rather the more local brain-body nexus realizes cognition? It is worth noting that cognitive media theorists who subscribe to 4E claims highlight the transformation of cognitive capacities through media (such as extended empathy in film; Smith, 2012; see also section "Seeing-in Pictures") and the centrality of embodied engagement (Nannicelli, 2019). Yet they mostly do so without assuming a literal extension of cognitive processes into the media artifacts we engage with, as EM would have it. According to such theorists, cognitive capacities and affective relations are still realized locally within an embodied agent.

The present paper does not focus on boundary definitions for cognitive systems. Also, specific media may pose additional challenges to an EM account for media interactions.[5] Yet I will return to some of those issues when discussing artifactual habits that are *locationally expansive.* There, I argue that a parsimonious theoretical assessment of certain media engagements captures organism and media as *jointly* exploring and bringing forth meaning or even models of the world. In any case, under the concept of habit otherwise seemingly disjointed processes (inner-organic and media processes) can be understood as unified (Fingerhut, 2020a). Although I do not focus on ontological claims regarding constitution and causality, I want to at least hint at a (in my view) promising way to challenge previous renderings of constitution. This could be accomplished by including dynamical and reciprocal causality between organisms and environment as part of what counts as constitution (Kirchhoff, 2015).[6] I generally agree with such accounts, which argue that the diachronic element of our history of engaging with artifacts (captured by the *temporal expansiveness* of habits) also has some bearing on the *locational expansiveness* of a mental state, as I will address below.

Enactivism and Domains of Value

As I argue in this paper, media engage us in an active exploration. A mainstay of enactivism is that perception and experience should primarily be understood as the activity of an organism. At the core of the enactivist approach (Varela et al., 1991; Thompson, 2007; Di Paolo and Thompson, 2014) lies the idea that we should understand all cognition as meaning-making against the backdrop of self-organized autonomous systems and their structured interactions with the environment they bring forth. Enactivism therefore unfolds around the concept of the metabolic organism and the autonomous self. It claims similar principles hold for single cells in their chemical environments, bodies-plus-tools in more evolved organisms, and human agents in social settings. The autonomous, living body is nonetheless at the heart of such accounts, which are organism-centered and that model cognitive activity in terms of its relevance to the viability of an organism. "Cognition, in its most general form, is sense-making—the adaptive regulation of states and interactions by an agent with respect to the consequences for the agent's own viability" (Di Paolo and Thompson, 2014, p. 76). Here, cognition is understood as a temporally extended dynamic and as an ongoing adaptive regulation.

The central cognitive activity is sense-making. This activity captures what we do when we bring forth meaning. Within such a concept, environment and organism can be seen as occupying co-constitutive roles (Thompson and Stapleton, 2009). The history of structural coupling between organism and environment leads to a form of convergence between the two, defining also what an organism is sensitive to in its environment.[7] Adaptivity is therefore also centrally interwoven with the sense-making of an autonomous system, in which it tracks whether environmental conditions are beneficial or detrimental for its viability (Di Paolo, 2005; Di Paolo and Thompson, 2014). The mutual co-determination of organism and environment occurs on evolutionary and ontogenetic timespans. But crucially, it is also is present in the immediate dynamics of the *here and now*. The latter is highlighted, for instance, in theories of "participatory sense-making" (De Jaegher and Di Paolo, 2007). Whereas some earlier accounts of autonomy-based enactivism focused on coupling with the environment mostly from the viewpoint of the organism, participatory sense-making gives socially negotiated cognition center stage—dispensing with the idea that relevant cognitive activity originates solely from a single organism.

Individual and interactive levels here are mutually enabling. Recent enactive accounts of language can be additionally seen as a media related extension of participatory sense-making. These accounts reference a central cultural domain within the human social niche: "linguistic sensitivities are the result of the specific contingencies and ecological co-constitution of our bodily existence in human worlds" (Cuffari et al., 2015, p. 1199). Yet despite this interest in the ecological constitution via "languaging" (Di Paolo et al., 2018), such accounts ignore

[5]The exploration of a scene in film, for example, relies on camera work and editing processes that have happened in the past. The perceiving subject is thus an active partaker in the succession of frames in the *here and now*, yet also a passive perceiver in terms of the many past operational decisions that they cannot influence. For discussion of pictorial artifacts in this respect, see Fingerhut (2014).

[6]Most EM theorists assume local, neural realizers when it comes to *conscious* mental states (and argue only for the extended nature of non-occurrent mental states, such as beliefs), whereas the dynamic-reciprocal accounts just mentioned also prominently address the unfolding of conscious experiences. Such accounts also encompass cultural phenomena (such as architectural contexts) that fall under the purview of what I call media (Kirchhoff and Kiverstein, 2019, 2020).

[7]By highlighting affordances for action within the environment, ecological psychology (Gibson, 1979) shares several tenets with enactive perception; the terminology of 'affordances' is thus used across theoretical boundaries. Ecological psychology comes with a set of further theoretical commitments that are not central to what I do in the present paper; I thus do not discuss overlaps and differences between ecological psychology and enactivism. For some recent discussions of this, see Ramstead et al. (2016), Crippen (2020), and Feiten (2020). See also section "Neuromediality and Media Affordances" below.

both the printed word and the use of language in other media systems as a factor in developing those linguistic sensitivities. Media (e.g., film, screen-based digital media, and printed words) entrain us in ways that are quite different compared to those of embodied, social languaging encounters 'in the wild.' But both the specific capabilities we develop with respect to these media along with ways language capabilities might transfer across media boundaries and into participatory sense-making constitute central questions for a media-informed enactivism.

This paper emphasizes the generation, sustaining, and active perception of values within an environment in structural coupling, by focusing on such coupling in media environments. Although the living body is a central reference point for enactivism, the enacted environmental loop it undergoes largely determines which mental state we entertain at a certain moment. That is what I will focus on by disclosing the sensorimotor and body-schematic dimensions of enactive sense-making in media contexts. As certain versions of what has been labeled *sensorimotor enactivism* argue, entertaining an auditory experience (to take just one example) differs from a visual experience based on the mastery regarding patterns of regularity between motor acts and sensory feedback (O'Regan and Noë, 2001). Both experiences differ, for example, from a thought in terms of the type of access to the world that they provide (Noë, 2009). From here, it is a small step to argue that media-sources engage us in media-specific loops with their own forms of access (Noë, 2012; Fingerhut, 2014).

Sensorimotor enactivism has been criticized for unnecessarily relying on (inner) knowledge with respect to the mastery of aforementioned regularities (Hutto, 2005). Despite this difference regarding knowledge, autopoietic and sensorimotor enactivism both agree that mental states cannot be fully captured by functional descriptions (of so-called knowledge obtained by the organism, or even in terms of the functional structures determining bodily loops through the environment). Enactivism could therefore be seen as highlighting the dynamic interactions with the environment more directly (Hutto and Myin, 2020). The specific unfoldings of such interactions is a central component of theories of participatory sense-making and has also been captured by the concept of "attunement" in enactive interpretations of skilled performance theories (Gallagher and Varga, 2020).

Enactivism claims additionally that the ability to generate and sustain values in our environment has to be part of a theory of cognition, proper. This also explains why enactivism relates to EM rather critically (Di Paolo, 2009). The functionalist descriptions that EM brings to bear in capturing mental states (e.g., our beliefs as brain-body-artifact hybrids) are based on the wrong model of the mind. It lacks reference to meaning-making—namely, to the body as a self-individuating system interacting with the environment (Di Paolo, 2009; Di Paolo and Thompson, 2014). Those differences can be unpacked in various ways. One main difference is that the continuous dynamic of regulating and adapting the body in sense-making also entails a concept of value and affectivity that other theories lack (Colombetti, 2017). Such values are sustained at different levels. These include the body in self-regulation, the body in sensorimotor coupling, and the body in intersubjective engagement (Thompson and Varela, 2001; Thompson, 2007; Di Paolo et al., 2018).[8] Cultural artifacts and media latch onto our bodies with respect to all three modes. For instance, clothing and the built environment alter our self-regulatory processes significantly by providing heat and shelter. Pictures and moving images engage us in a sensorimotor coupling that differs from engagement with depicted scenes in the flesh. They thereby enable us to attribute a different system of values to those scenes. Digital media, in turn, constantly alter our social interactions. Generally, by co-constituting domains of interaction, media embody meaning. This is because they have become part and parcel of the strategies by which the human body engages the world.

While functionalist descriptions cannot fully account for the generation and sustaining of values, I would argue *pace* enactivism that when it comes to the tracking of such values, neuronal mechanisms and bodily sensitivities that enable such tracking constitute a central level of description.[9] Cognitive neuroscience might therefore capture how our visual system interlocks in perceptual engagements with certain artifacts (how, e.g., a film entrains us). Theories of emotions, in particular, might explain how specific emotional states track values in our environment based on embodied profiles that afford specific kinds of cognitive processing (Prinz, 2004; Fingerhut and Prinz, 2020, forthcoming). When it comes to cultural domains and media, one should think, moreover, of regulatory principles and norms for our bodies to sustain that go beyond avoiding harm and satisfying the need for food or shelter. Our bodies, for instance, might be seen as exhibiting a need for information and exploration. This is exemplified in the affective states of interest and curiosity, which might explain the pleasure we take in a wide variety of domains including media (Biederman and Vessel, 2006). We might therefore also think of further affective and aesthetic engagements that media afford, such as wonder and play, through which we track what we value in the arts (Fingerhut and Prinz, 2018).

Artifactual Habits

Enactivism argues that we bring forth experiences by engaging with the world and others. Such active engagements differ substantially when we engage with a social scene in a film or explore the world as it is depicted in a photograph. Different pervasive artifacts and media contexts might also have led to the emergence of different bodies (or body-schematic processes) that we bring to bear in such media ecologies. Walking through the built environment of a city, for instance, requires a set of bodily

[8] Although it could be argued that the *sensorimotor enactivism* of O'Regan and Noë (2001) does not centrally capture this reference to affective states and the autonomous body in need of coupling (Fingerhut, 2012).

[9] In this sense I would argue that it remains explanatory necessary to identify specific structures in the brain-body-world nexus (i.e., in artifacts and human bodies) that jointly realize those loops, while at the same time retaining the possibility that cognition in the relational sense might have no location proper (see for an excellent critical discussion of this: Walter, 2014).

engagements different from the one we employ when seeing a movie in a cinema setting.

By directing attention toward what can be called 'artifactual habits of exploration,' I aim to capture the salient differences between those situations. This paper argues that human cognizers are constituted by a plurality of habits that bring forth their own domains of interactions and respective ranges of viability. Habits are structured ways of acting and central loci of meaning-making. It is only when something has either entered into a pattern or is registered as a violation of such a pattern that it becomes meaningful to an organism. The rest is noise. As pragmatist philosophy in particular has acknowledged, habits can therefore be seen as the basic building blocks of the mind: "the medium of habit filters all the material that reaches our perception and thought" (Dewey, 1983, p. 26).[10]

For the present paper, it is central that certain habits can be described as mixed media affairs between bodies and artifacts. This links them to the debate regarding the extended mind and they therefore can be captured by one meaning of *expansiveness* identified earlier. The bodily interaction that pertains to a habit is co-determined by the media artifact. In other words, the engagement unfolds according to media-specific processes. The habit is then re-instantiated each time the brain-body-media coalition is formed. Habits are *locationally expansive* in this sense and in their reliance on external structures of the designed environment, of cultural artifacts, and of media more generally.

Habits also share the quality of being *temporally expansive*. This means they bring our history of environmental coupling to the *here and now*. They thus structure our actions and determine our tacit expectations with respect to a domain (Fingerhut, 2020a). In many ways, habits are comparable to skills. For the purposes of this paper, habits and skills largely function as interchangeable concepts. But in contrast to skills (Fridland, 2017; Hipólito et al., 2020), habits do not require the same level of control in their development. Moreover, they can be acquired and molded simply through exposure and implicit statistical learning. The temporal expansiveness of habits nonetheless exceeds any concept of repetition: "rather than being the repetition of action, habit is characterized as the open and adaptive way in which the body learns to cope with familiar situations" (Miyahara et al., 2020, p. 125).

Habits are not merely rigid mechanical routines. Rather, they constitute flexible ways of world-making and capture how human cognition may be cultural *tout court*: cultural contexts, artifacts, and media latch on to existing modes of perceiving and affective engagement, moving them toward new forms. As such, artifactual habits constitute an interactive domain between organism and environment. Given this, they are determined as much by external media as they are by the activities of the organism. This relates to the third aspect of expansiveness. Artifactual or media habits are proven to be *transformatively expansive*: they generate new patterns of interactions and domains of value in the process of reciprocal adaptation between organism and cultural environment. Some propensity to pick up and integrate new patterns must obtain on the side of the organism (i.e., as an enabling condition), yet artifacts, media, social environments play the more active role in driving such transformations. Technical innovations force us to learn new skills; statistical immersion within new (typically urban) environments or new social media may alter our habits of interpersonal engagement; and finally, cultural innovations and especially the arts may challenge our habits of engagement in various respects.

The account of habits proposed here portrays us as expert performers in different media settings. Synthetic accounts of skilled performance have already addressed some of the competences this entails, along with the flexibility of habits I envision, for other domains (Christensen et al., 2016). For example, Gallagher and Varga (2020) describe a horizontal axis involved in the joint performance of music. This axis stands in opposition to a vertical one involving higher cognitive processes interacting with bodily engagements. The horizontal axis includes processes that "extend into the world, meshed with the structures of our intercorporal and material engagements" (Gallagher and Varga, 2020, p. 7). This is *locational expansiveness*, to use my term. Understanding such attunements and the dynamic, situated processes in performance studies (but also in media context in which we turn out to be expert performers with respect to media artifacts) could centrally inform our understanding of situated cognition as those authors argue.

I discuss examples of media engagements more extensively below, when I put the account of artifactual habits to work (see section "Toward a New Cognitive Media Theory"). But to get an idea, consider cinema. Edited Hollywood movies rely on us exploring their content according to medium-specific patterns. Some of these include specific camera and lens movements or editing techniques that could involve switching perspectives to portray a scene, or a montage to exemplify an idea. Movies are designed by employing film techniques that have evolved over time. Some of these techniques instill immersion in us viewers, which seems to be a central aim of Hollywood cinema, and engage us with configurations that entrain us with their content (a situation, a scene) in specific sensorimotor or affective ways. Despite feeling immersed in such situations it should be clear that these engagements differ significantly from how we could experience a situation or scene in the flesh. We might not be aware of this anymore, but film is contingent upon on us having integrated certain techniques of exploration into our habits of seeing.

With respect to film (as opposed to static images or written text), it is interesting how some of the activity of exploration sides with the medium itself. Film theorists have aimed to capture the ways we lend our body to the medium in such cases. In the process of doing so, it has been argued that we engage a "surrogate body" (Voss, 2011). One way to capture the embodied engagement in these cases is by exploring a specific "filmic body schema" that extends into the filmic realm and expresses itself by engendering certain film-specific embodied engagements

[10]For an excellent overview on current pragmatist theorizing on habits, see Caruana and Testa (2020).

(Fingerhut and Heimann, 2017). Some initial thoughts might help demonstrate the plausibility of such a concept. In film viewing, our self-initiated, real-world related movements are attenuated in ways that free up resources for an intensified engagement with the cinematic works themselves experienced as bodily engagement (i.e., with camera movements, editing, perspectival change, as I will address in more detail below, section "The Filmic Body Schema").

THE ROLE OF THE BRAIN IN THE MEDIA MIX

Neuromediality and Media Affordances

Above, I alluded to radical predictive processing (RPP). This perspective weaves "designer environments" into a novel way to understand the brain (Clark, 2015a, 2016). Predictive processing theories generally agree that the central function of the brain is to adjust the organism to its environment by using multileveled probabilistic predictions. In RPP such inner models are seen as action-oriented through and through. They have the function to enable an efficient, and highly context-sensitive grip on structures and scaffoldings in the environment by making "use of multiple, fast, efficient, environmentally-exploitative, routes to action, and response" (Clark, 2015a, p. 18).

In media ecologies, this grip takes on a specific, even more interlocked nature, because media, among other things, have been designed to engage and entrain us. Before going into some of the details of such media engagements, it might be helpful to account more generally for the contribution of the brain in embodied media interactions by introducing the concept of *neuromediality*. Such a concept aims to relate neural activity to artifactual habits of perceiving. By highlighting processes that correspond directly to media engagements, we can avoid falling into a bio-, or socio-essentialism. Such essentialism treats media as something that only impinges on a cognitive system, which itself has evolved and developed in our every-day interactions (e.g., either face to face with others or in the exposure to natural objects) and interprets neural data in this way. Under the proposed neuromedial perspective, neural responses can also be seen as being exapted for media contexts. One aim is therefore to identify neuronal contributions to new dimensions of interaction that cultural artifacts, such as pictures and moving images, afford. The pervasiveness of such media can be speculatively related to the impact of other human artifacts on the brain, which has been explored with respect to the organizational principle of "neural reuse" that has been mostly explicated in relation to tool use and language processing (Anderson, 2010; D'Errico and Colagè, 2018). To date, there are no comparably sophisticated accounts for artifacts beyond language (such as depictions, which arguably occupy a longstanding and central role in human cultures, Brumm et al., 2021).

Notwithstanding such accounts of how neural circuitry integrates new functions, it is generally important for a cognitive science of media to build upon some normal conditions of media-engagement that have developed ontogenetically through experience-based learning and statistical immersion. This is true not simply with respect to images, but also for film and TV, the built environment, and digital media. Such considerations will be instrumental in developing a theory of how artifactual habits differ from each other and how an artifactual habit finds expression in a specific media ecology or cultural environment. They can also help map combinations of media components and the neural-bodily resources on the organism side that they draw on. In a second step, this approach can then address the question of how the quality and content of an experience is determined by habitual patterns of engagement (and the deviations from the norms those habits track)—and how we enact a specific picture, film, or novel.

What do we actually perceive when we engage with media? As I argue, media provide models of the world that a cognizer can latch onto in media-specific ways. Artifactual habits describe such ways of enacting models. Yet it is not the model itself that shows up in our consciousness. Instead, we perceive certain scenes in the forms that pertain to different media (e.g., in pictures, films, and novels), we engage with utterances of other people (e.g., in social media), or we perceive opportunities to move (e.g., in the built environment).

This relates to an understanding of our perceptual system as geared to pick up opportunities to act, which is explored in ecological psychology (see footnote 7). Concepts such as "affordances 2.0" neatly capture how those opportunities to act change dynamically in human-environment systems (Chemero, 2009, pp. 150–4). Here, environmental affordances for action are not just properties available for pick-up to a pre-existing body with specific sense organs (Gibson, 1979). Instead, cultural niche and sensorimotor capabilities are constantly altered on short timescales by human animals acting in these niches. It is in this dynamic sense that affordances have also become a central concept within recent theorizing about the cultural environment and the enticements it contains (Withagen et al., 2012; Rietveld and Kiverstein, 2014).[11]

With respect to different media, then, one could argue that affordances correspond to habits or skills that are the topic of this paper. These central, media-related affordances have to be theoretically modeled in terms of the media-related habits that correspond to them. For example, a depicted door is perceived as walk-through-able in a way that is different from a door in a building. Insofar as media expand our sensory system and co-structure our habits of perception, they also generate new affordances. This pertains to how affordances differ systematically *across* media habits (e.g., the differences between watching a movie, reading a text, or engaging in a social media chat). Another question is how affordances are dynamically modulated *within* a media engagement. The concept of 'interaction-dominant dynamics' describes one such dynamic between media artifacts and the brain-body nexus—one that captures how an explorative activity is guided by a media ecology. It has been argued, for instance, that the mouse-computer system entrains the user into a certain pattern of

[11] For sociocultural affordances in social relations, see Ramstead et al. (2016). For social affordances in digital media, see Fox and McEwan (2017). For affordances in architecture, see Jelić et al. (2016) and Djebbara et al. (2021).

action (Dotov et al., 2010, p. 3). In such cases, neural activity is modulated by the sensorimotor-artifact dynamics of the larger system. This includes switches in processing that could enable the peripersonal space (Làdavas, 2002) of the engaging organism to extend into the virtual environment of the computer screen (Bassolino et al., 2010). After such a switch, the receptive field of certain neurons changes significantly. Objects within the virtual space take on a different presence and the organism engages in a different cognitive processing style. Such kinds of entrainments might even be more intense in new media devices such as virtual reality (VR), where they are used for motor-cognitive neurorehabilitation (Perez-Marcos et al., 2018), yet they can be traced for other media as well.

The point I want to make is of a general nature: understanding different media requires a focus on how media structure our engagement with the worlds we are presented with. We need a view of the brain as sustaining a dynamic and flexible neuro-cognitive architecture (i.e., one that switches between and locks into different media). Here, as before, I suggest the utility of the concept of neuromedial processes for denoting the contribution of the brain in such dynamics without giving it exclusive importance in defining the structure of the relationship to mediated worlds. The way certain media store, process, and transmit information makes them specific model-environments that pre-structure such relations for the human organism. It is—or should be—the task of an enactive theory of media to highlight how we attune to such models and what we can do within them.

The Radically Predictive Brain

I suggest capturing the dynamic and flexible cognitive architecture in media engagements by philosophical predictive modeling accounts. Clark's action-oriented version, labeled radical predictive processing (RPP), focuses on the role of the brain in recruiting resources for action (Clark, 2013, 2015a,b, 2016). He provides a theory of the neural system as engaging in active self-organizing dynamics that also could make salient how the active body becomes recruited by designer environments that themselves constitute central models for our mind.

The general idea of predictive coding (PC) is that in terms of perception, cognition, and action, the computational contribution of the brain involves providing a multilayered system that produces predictions or hypotheses about the world. The brain reduces uncertainty about its environment by engaging in "prediction error minimization" (Friston and Kiebel, 2009; Friston, 2010; Friston et al., 2010). The theory assumes that predictions cascade in top-down flows, from higher layers toward lower ones. They are met by upcoming flows of information that either match those predictions or not. The brain deals with incoming information in a cost-efficient way by propagating residual prediction errors in the system (rather than construing a representation based on sensory input).

Predictive coding theories assume that the brain became wired to run an inherently culture-dependent model of the world that controls the body in cultural ecologies through predictive processes (Gendron et al., 2020). Enactivists criticize such predictive theories for their reliance on inner models or 'priors' as hypotheses. For them, these bear too much resemblance to inner representations as central explanatory elements (Hutto, 2018; Hutto et al., 2020). Clark (2015a) sees his radical version as being fit to oppose such a criticism, because it treats the brain as mainly engaging dynamical loops through the environment (with the external designer environments constraining these loops, more on this in a bit). He claims that RPP further alleviates explanatory weight from inner generative models (that remain a central element in his theory) by spreading this weight onto the ongoing interactions and the environmental structures themselves.[12] Along those lines it has been emphasized that one way to reduce prediction error is to test the environment by actively engaging with it, which falls under the concept of 'active inference.' Here, the motor system can be described as part of cognition in oculomotor control (for example) as well as in cued and goal-directed movements (Friston et al., 2010; Adams et al., 2013; Constant et al., 2020a).

Clark's (2013) concept of designer environments directly focuses on how material culture structures our intersubjective take on the world. Public symbols are effectively forcing upon us new regimes of pre-structured, re-entrant information processing.

> The same potent processing regimes, now targeting these brand new types of statistically pregnant designer inputs, are then enabled to discover and refine new generative models, latching onto (and at times actively creating) ever more abstract structure in the world. Action and perception thus work together to reduce prediction error against the more slowly evolving backdrop of a culturally distributed process that spawns a succession of designer environments. (Clark, 2013, p. 195).

Clark mostly discusses lingua-form perceptuals that are public, external models of the world (such as language, formula, theories; Lupyan and Clark, 2015). Still, such a view can include media, cultural artifacts, and the larger cultural environment to support claims regarding artifact engagements (Constant et al., 2020a,b). In the quoted passage, Clark's focus is on cognition and thought. By emphasizing how the structured environment contributes to cognition, he aims to appease the worry that predictive processing does not provide enough internal structure to explain our full-blown cognitive architecture. Yet, what he claims for the "abstract structures in the world" I would argue also applies to the experimental regimes that media present to us. Clark even references different media and their material properties that limit our interaction space (e.g., computer-keyboard interfaces and specific video formats) that

[12]Clark argues that the actively inferencing organism is not decoupled from the environment. It constantly updates its predictions or priors in a way that they no longer resemble classical mental representations anymore (Clark, 2015b). Others argue that an enactive account of predictive engagement (PE) should further do away with inferences and models in the theory. Instead, it should directly focus on the situation dynamics of the whole system along with concepts such as "adjustment, attunement, and accommodation" (Gallagher and Allen, 2018). I am greatly sympathetic to their version of predictive engagement, but do not see it in strong opposition to my understanding of the RPP account presented above. As part of my survey of 4E and related accounts, I have chosen to focus on RPP because it is more directly geared to an understanding of media as designer environments and could be seen as an extension of the extended mind views developed earlier. In contrast, Gallagher and Allen (2018) focus on the dynamics of social interaction.

are nonetheless key to or cultural ecosystems (Clark, 2016, p. 279–281). In this, they are a central part of the ever-faster succession of designer environments. Media entrain our perception-action cycles. Despite and precisely because they thereby reduce the complexity of (embodied) interactions with our surroundings they also enable us to engage in new and potentially exciting explorations (as we will see with respect to media works such as texts, films, etc.).

The PC framework sees perception as largely operating based on generative models (conditioned probabilities that link data to their hidden causes in the environment) in a top-down way. These operations start with the inward layers of a hierarchical model of the brain. RPP shares this basic assumption, but it enables us to include cultural environments as part of the predictions more systematically. The way the brain dovetails with designer environments could render these environments an outer layer of predictions themselves, generating their own media-specific flow of information. One might still worry that a separation of inner and outer processing is re-introduced, rendering the environments as passive contributors to the inner complex and active machinery. In this scenario, they would function simply as input to the cognitive system.[13] Another worry is that the 'free-energy minimization' that is part of the larger theory unifying biology and cognitive science introduces an overgeneralization that contains the assumption that a system should seek out states and therefore environments that would contain no surprise (known as the "dark room problem," Friston et al., 2012). Media environments seem to present the opposite of this. Although I do not believe that RPP can fully deal with those worries on its own (for this the larger, more enactive picture form above would be needed), I nonetheless will address some answers from within the framework, because this also helps to see more clearly how media environments could fit into the predictive picture.

Active Media Inference

The first worry is that designer environments still seem separated from making a central contribution to cognition. Neural processing of generative models in hierarchical layers of the brain supposedly does most of the work. This worry can be partially assuaged by pointing to the role of action within the active inference concept in RPP and the targeting of different layers of generative models. As we have seen, a central way to reduce uncertainty is to act upon the environment. This allows for an enhanced hypothesis testing. Such a picture is alluring because it can also capture the ways our actions in active inference are pre-structured and limited in designer environments (and media ecologies). It simultaneously addresses how the dovetailing of brain-organism-artifact via this pre-structuring facilitates the organism in engaging with the richness and potency of ecological information.[14]

Once again, consider our brain at the movies and the case of perception. Here, the visual system's priors are not neutral between many possibilities to engage. Rather, they operate within a limited range of possibilities. In typical Hollywood cinema, for example, we do not have to explore the scene presented on our own: the director, camerawoman, and editor all direct our attention to the salient part of the action. Our eye- and head movements are thus cued (Loschky et al., 2015). In such cases, activity independent of such cues (e.g., saccades to different areas of the screen) would not be rewarded with the relevant information that drives the story. Certain actions, such as standing up and moving toward the screen, won't yield relevant visual feedback. Seeing to people engage in a movie scene can thus be contrasted with perceiving a scene wherein two people engage in the flesh. Once we have switched to the regime of film (i.e., reduced uncertainty with respect to the more global environment; enabling a specific set of generative models and hyperpriors), we allocate resources to other elements we would not necessarily focus on in real life (e.g., by enhancing our emotional engagement in the close-up of a face). In this scenario, active inference based on sensorimotor filmic priors allow us to engage with an idea, character, and story in ways that would not be available in the real world, especially because certain actions within such a media ecology are reduced and others are taken over by the medium (e.g., by zooming into a scene). Film therefore constitutes its own generative (cause-effect) model. Here, the presence of a medium that adheres to certain regularities in conjunction with layers of neurons engaged in the minimization of prediction error jointly manage the kind of sensory flow within a media habit.

The degree of alignment with an environment that I just described is, for example, captured by variations of "precision weighing" that modulate the impact of error signals in specific contexts (Clark, 2016, pp. 57–59). Precision weighing provides a mechanism that plays a role in what we pay attention to Feldman and Friston (2010); Parr and Friston (2017)—one that has been employed to understanding "presence" in both media and non-media contexts (Parola et al., 2016; Seth, 2019). Take another example. Walking through a built environment (such as an apartment, university, or city) renders certain kinds of information more or less salient. This leads to greater precision, and therefore less uncertainty, in embodied predictions about certain elements. This is, for instance, expressed in a high conditioned probability the streets in a city follow a grid-like structure. Violations within such a geared prediction regime will gain our attention more easily. RPP therefore provides an organism-artifact mixed-media model that, in the end, could be part of an explanation about why certain forms of attention or affective engagement, etc. occur within a specific habit but can be quite different in another media environment. Moreover, the structure of the designed media environments co-constitutes our engagements with generative models in the brain being geared to pick up and integrate recurring patterns.

[13]The formal description of systems that engages with active inference (i.e., described within the boundaries of a Markov Blanket) could also include elements outside the living organism. In this sense, it would be an outer layer of a nested system (Kirchhoff et al., 2018). But without further explanation, such an outer layer would still seem to remain at the periphery of what constitutes cognitive engagement. What I try to argue is that we attune to external models at different levels of our hierarchical generative model on the organismic side.

[14]But see Anderson and Chemero (2018).

Culture as the Plurality of Mutual Models

Designer environments are thus centrally involved in eliciting switches between generative models in the brain (or what could be considered hyperpriors, such as when switching between perceiving a picture and a social scene *in the flesh*). Even more centrally, however, the external models co-determine the ways in which multilevel, probabilistic models unfold deeply within the engine of the human cognitive system. This view of the cognitive system can therefore do without assuming that we have to represent the structures of the media artifacts themselves. Instead, the brain-body nexus jointly with the medium engages in exploration. The first worry, that of a secondary contribution of media-designs, is thus addressed to some extent. Still, the second worry remains, namely that our engagement with "statistically pregnant" designer environments does not seem to fit the general aim of organisms to reduce uncertainty.

Regarding this second worry, I would like to steer clear from discussions of a dark room that immediately presents itself as an adaptively unreasonable and unsuccessful coping strategy that leaves seekers of dark rooms at an evolutionary disadvantage (it remains problematic that the theory might proposition such a scenario). When it comes to artifacts and media, the more relevant discussion is the perceived value of experiential surprise (Van de Cruys and Wagemans, 2011; Seth, 2019). Predictive Theories based on free-energy minimization do not seem to account for the "deep, positive attractions of novelty, play, and exploration" (Clark, 2018, p. 524). Clark discusses this in terms of an "information theoretic subversion," which is the idea that we could describe a predictive system maximizing prediction success (avoiding the dark room) and still end up with a perfectly trivial sense in which the system achieves that. Such subversions seem to be forestalled by the plurality and dynamics of our cultural practices, artifacts, and media.[15] They come to us with new affordances for engagement, with a multitude of complex traditions ready for exploration, and by implicating novel epistemic actions.[16] Such designer environments thereby ensure "a steady diet of change, innovation, and challenge" (Clark, 2018, p. 531).

This speaks directly to the aforementioned paradoxical aspect of habits as sustaining certain ways of acting while, at the same time, evolving to incorporate new forms of engagement (being *transformatively expansive*). Habits seem to minimize novelty by attuning us to a specific designer environments or media settings. They are therefore conservative in the sense of providing and keeping us within a range of viable actions. Yet since habits are partially constituted by the pervasive artifacts that evolve around us (they are *locationally expansive* in that media co-constitute their exploration), they also can appear as more progressive.[17] We are exposed to a plurality of designer environments that we co-construe and that still dynamically evolve. In engaging those environments, our inner models and the outer models coalesce. What is more, they become mutual models that span brain, body, and environment, that are actively embodied, and which are shared with others.

These are only cursory remarks. Still, RPP provides an initial theory of how the brain folds media environments into our expansive sense-making activities (with the caveat that it still relies on inner models in ways enactive theorizing would object to, see footnote 12). It claims that the brain-body system picks and engages strategies for dealing with the world based on error minimization and active inference. Media environments, in turn, provide a plurality of strategies for dealing with the world via experiential models, models that constitute the shared space of culture and innovation.

A SHORT PRIMER ON MEDIA THEORY: THE MEDIUM IS THE MESSAGE

The current paper proposes understanding enculturation by employing a theory about our embodied habits in relation to external media. Here, habits are media-inclusive, temporally outreaching, and governors of the dynamics of our engagement. The premise is that media widen our senses and are central conveyors of culture. Before I discuss how this account of artifactual habits helps us tackle specific media engagements (see section "Toward a New Cognitive Media Theory"), it is worth taking a quick detour to see whether the central tenets of situated cognition relate to a more general media theory.[18]

A seminal position within the admittingly diverse field of media theory is McLuhan's media ecology (McLuhan, 1962) that still promises to evolve into exciting new directions (Lum, 2014). Media ecology probes the effects of anything we use in dealing with the world around us. For instance, McLuhan even includes lightbulbs as media. He does not focus solely on mass communication, but on how media enable us to do things. By his definition, media are extensions of the human body. They span bodily functions ranging from basic needs to cognition. This explains why McLuhan's concept of media as "extensions of man" includes housing and cities as extensions of bodily heat control (McLuhan, 1964). This is obviously in addition to more classical areas he touches upon such as TV and movies (which extend our sensorimotor grasp) as well as the now-ubiquitous electronic media that are seen as an extension of the human nervous system (McLuhan, 1964, 1988).

[15] Although, such subversion could be attributed to certain domains of our digital media environment. Consider the rise of casual puzzle games such as Candy Crush and Gardenscapes, which achieved 180 million downloads by 2018 (Katkoff, 2019). Such games present players with successive puzzles of ever-so-slightly increasing complexity.

[16] For the concept of epistemic action see Kirsh and Maglio (1994) and the discussion in Clark and Chalmers (1998).

[17] Habits evolve and find new expressions in the succession of media forms (reading, e.g., transitioned while its medium changed from handwritten texts, to printed books, to current tablets devices) or they emerge as new habits of engagement as in the case of more radical technical or artistic innovations (think, again, of moving image devices).

[18] I do not aim to capture the multi-faceted field of media studies, the scope of which goes well beyond this paper. One reason is because many accounts in media studies combine cultural analysis and the philosophy of technology with normative claims. This includes reflections on the tyranny of digital media and computational thinking (Stiegler, 2019), the implications of the "neuro-image" in socio-political terms (Pisters, 2017), and the aforementioned critical assessment of the attention economy (Crogan and Kinsley, 2012).

Three things are relevant here. First, one of the more established distinctions in the amorphous field of media theory is its relative separation from communication theory. The latter predominantly focuses on the sender and the receiver, the source, and the destination of messages. Where communication theory describes what part of the message gets through (treating disturbances in the media channel as noise), media theory aims more directly at the media qualities of the given channel and the way external devices record, process, and convey information. Versions of communication theory based on Shannon and Weaver's (1949) information model already had their impact on philosophy, such as in terms of the naturalization of intentionality in representational theories of mind (Dretske, 1981; Adams, 2003). It stands to reason that media theory could play a similar role within 4E cognition. Understanding the mind requires more than a focus on what information gets *in*. This understanding has to explain how mental states are brought forth in embodied engagements that are based on the cognitive practices I have described as joint explorations of media and organisms.

Second, media theory provides a way to centrally understand *culture* that spans technology and images, social engagement and art (Bickenbach, 2011). At the same time, it captures the decisive impact media have on the mind and the human sensorium (Gane and Sale, 2007; Jones, 2010). In this, it complements reconstructive evolutionary accounts of culture as social learning in biology (Heyes, 2020) and cognitive neuroscience (Gendron et al., 2020) by focusing on the aspects of learning and adaptation that are mediated by media. The humanities background for media theory could supply additional help in tracking the concept of value or significance across different disciplines, while also challenging conventional ways of thinking in the cognitive sciences. As an enactive category, artifactual habits involve more than just habituation. They decisively encompass a capacity to generate, sustain, and track values in the environment. Enculturation can then be understood as an extension of such value systems: "culture thus concerns all forms of significance that are common to groups of people and inherited by social rather than genetic means" (Durt et al., 2017, p. 74). 4E-supported media studies could explore enculturation by not focusing solely on social interactions with others (Veissière et al., 2019): it could instead achieve this by foregrounding the cultural artifacts and media domains that centrally permeate and structure our minds.

A third point, frequently made in media studies, is the claim the impact of media is so pervasive and ubiquitous, their co-constitutional role for our (cognitive) lives does not come to the fore anymore. As Bourdieu (1977) developed with respect to the concept of *doxa* (as opposed to the more explicit *dogmas* and norms in a society), culture could be seen as all the things that are taken for granted in a society. A theoretic effort is required to make explicit the ways in which we are enculturated. The reign media have over us is one that relates to their structural impact. This is captured in McLuhan's most famous phrase: "the medium is the message" (McLuhan, 1964). In the sense of information or content, no message can measure up to the effects of the structural interaction enabled by the medium that carries the content. This makes McLuhan's observation a theoretical call to the arms—one that extends to philosophy of mind that might be prone to miss out on the potentially profound impacts of media. It is therefore important to include a wide range of media and cultural artifacts to understand this impact (as I do in the next sections). While their impacts may not always be immediately transparent, they nonetheless form an infrastructural basis for experience and understanding.

TOWARD A NEW COGNITIVE MEDIA THEORY

We saw that within a situated cognition perspective, some tenets of a general media theory could also constitute tenets for a philosophy of mind. Despite case studies in specific domains such as the internet (Halpin et al., 2010; Smart et al., 2017; Clowes, 2019), attempts to include a more general media theory within situated cognition are sparse.[19] In media theory, *cognitive media theory* is most directly related to questions regarding the kind of mental states we entertain in our media engagements. These range from story engagements to aesthetic evaluations (Nannicelli and Taberham, 2014). For the remainder of this paper, I explore some media domains under its auspices. With this exploration, I intend to put the proposed artifactual habits account to work.

The Filmic Body Schema

In the 1980s, film studies took a naturalistic turn that challenged the prevailing Big Theories of its time. The turn drew more systematically on research from linguistics, anthropology, evolutionary biology, psychology and neuroscience (Bordwell, 2013). The so-called 'cognitive media theory' claimed that the widespread impact of cinema "must be connected to some fairly generic features of human organisms to account for their power across class, cultural, and educational boundaries. The structures of perception and cognition are primary examples of fairly generic features of humans" (Carroll, 1985, p. 92). Filmmakers achieve their effects by eliciting emotions and guiding our attention by story and character development—but also by framing, camerawork, and editing. In this respect, movies are *attentional engines* (Carroll and Seeley, 2013; Seeley, 2020). Cognitive film theory never explicitly stated that it is committed to a basic set of cognitive mechanisms. It nonetheless rests on a fixed-properties view of the mind that the present paper wants to challenge by providing a more integrative and dynamic theory regarding our cognitive capacities.

An often-reported finding is the amount of viewer synchrony during feature films. Through an inter-subject correlation analysis of fMRI data from participants watching a movie (*The Good, the Bad, and the Ugly*), Hasson et al. (2008) found an exceedingly high convergence of activity. As other studies have confirmed, such convergence is higher for edited film clips compared to unedited ones (Herbec et al., 2015). This could support the universalist claim of cognitive media theory because

[19] An exception is Logan (2013).

it appears to establish the existence of generic features of the human cognitive system that cinema plays to. Edited sequences entrain us in their unfolding more than non-edited ones, as do moving images more so than static pictures. The latter claim been demonstrated with respect to "attentional synchrony" using eye-tracking paradigms: compared to static scenes, sequences with actions and movement generate greater attentional synchrony, with respect to fixations and saccades in participants—especially when tracking people and faces (Smith and Mital, 2013).

The general attentional synchrony for dynamic scenes has indeed been exploited by film to hide its media features (e.g., camera movements or editing). Particularly for Hollywood cinema, montage adheres to what has been labeled 'continuity editing'; these are shooting and editing rules aimed at creating smooth, visual continuity in the eye of the beholder (Berliner and Cohen, 2011). The rules include perspectives and camera angles that can be assembled together before and after a cut (for instance, one should remain within an angle of 180 degrees and not go below 30 degrees). Often, the movement of an object or person is preserved when there is a cut. This ensures that such "match-action" cuts keep us entrained (Smith and Martin-Portugues Santacreu, 2017). By employing these techniques, there is a high propensity that our engagement with medium-specific characteristics such as edits does not reach conscious awareness anymore (Fingerhut, 2020b), or, at the very least, are subdued. This is captured by a phenomenon called 'edit blindness.' 30 percent or more of cuts go unrecognized within a scene, even when the viewer is tasked solely with reporting cuts in 5-min clips from Hollywood blockbusters (Smith and Henderson, 2008).

From the perceptual cognitive neuroscience perspective, each cut constitutes a significant event or violation of expectations. The neural signature of a film cut resembles that of a syntactic violation in language processing or in the order of sequence for comic-like stories using static images (Magliano and Zacks, 2011; Maffongelli et al., 2015). Let's return to the cuts described above. When comparing continuity edits to those that depart from the rules, no significant differences in early visual or syntactic processing were found. Instead, differences appear in brain areas that process violation repair. In cases where such post-perceptual updating is not occurring, other areas (such as those related to the conscious processing in detection tasks) are found to have neural signatures resembling those when a change is detected in a change blindness paradigm (Heimann et al., 2017).

One interpretation of these findings is that the visual entrainment to depicted elements (perhaps the movement of an object or person before and after a continuity edit) is sufficient to suppress conscious processing of cuts, allowing viewers to engage with the scene. In non-continuity editing, those content-related cues are simply insufficient to suppress awareness of the filmic means.

Yet one could also argue that continuity editing only works because it is integrated into a learned habit of enacting film. This would mean editing recedes into the background (and escapes our attention) only after we have developed a pictorial, moving-image competence. First, we must have had some exposure to edited film. Only once we have incorporated our filmic explorations (through camera and editing) into an artifactual habit of perceiving, may we stop perceiving these discrete configurational elements as independent elements, or events. Indeed, there is some experimental support for such a view. First-time viewers of film do have trouble perceiving spatiotemporal continuity in a scene that is put together adhering classic editing rules. Due to cuts and perspectival changes, such viewers do not perceive what is depicted before and after the cut as one and the same object (Ildirar and Schwan, 2015; Ildirar and Ewing, 2018). One explanation for this is that first-time viewers perceive cuts as a strong distortion—not just as a perspectival shift displaying the same scene. The flipside of this is experienced viewers of film have integrated such violations as part of film viewing and have developed a filmic habit of engagement. This then can be seen as one element of a filmic habit that comes with its own sensorimotor rules or even body schema (Fingerhut and Heimann, 2017). And it is only *within* such a filmic body schema that we can explain how attention and emotions are employed while experiencing a story in a way that captures what makes our engagement special in such cases.[20]

The present paper assembles phenomena from different media domains, thereby exploring how best our cognitive engagement may be described. This includes focusing on how external media and neural processing should be combined in terms of the realization base of mental states as well as focusing on enactive sense-making and the habits that structure such sense-making in different media ecologies (with habits constituting a central level of description in 4E media theory). I furthermore argue that predictive theories could fit neatly into this picture, for they can explain how we engage with media works (e.g., what predictions we bring to bear when we, for instance, watch a melodrama, a TV crime series, a horror movie, or read a novel). The more radical version of predictive processing discussed may additionally capture how we share into the explorative world-models designer environments present us, rendering them mutual models.

I am not aware on any substantive work on predictive coding and film. Nevertheless, there are interesting attempts to apply predictive models to works of literature, namely by treating literary texts as *probability designs* (Kukkonen, 2014, 2020). Generally, Kukkonen argues that literature engages in enhanced interoceptive explorations, referring to claims that inferences in hierarchical PC encompass exteroceptive and interoceptive prediction errors alike (Seth, 2013). Since the medium (in this instance, texts) limits our range of actions within a media environment, it makes specific elements more salient, allowing us to further explore affective evaluations of our inner realms that might otherwise go unrealized. Predictions here unfold on several levels, the most important one addressing narrative and plot. Given my focus on sensorimotor and body-schematic processes, I am more interested in the embodied reading experience and the designed sensory flow in such engagements. Here, *form* emerges as the central concept. Form, which is "foregrounded in the designed sensory flow of the sentences[,] sparks epistemic active inference, but arguably [it] also [serves] as [an anchor]

[20]While continuity editing therefore holds some interest to film studies, it also is too limited in its purview. Film scholars aim instead to understand how editing mediates the emotional and Gestalt perception within our filmic habit of engaging (Pearlman, 2017).

in the text to return to" (Kukkonen, 2020, p. 189). Because real-world bodily engagement is attenuated in reading, literary structures and formal elements can channel sensory flow in media-specific ways. On this point, compare how, in film, both editing and camera work scaffold our immersion and determine our engagement. However, the ability to return to those specific anchors earlier in the film experience is largely precluded. Therefore our self-initiated embodied engagement might be even more reduced compared to literature (in which we could, e.g., saccade or scroll back to earlier passages in the text). In engaging a filmic body schema under cinematic conditions we surrender our motor activity to the medium. This therefore constitutes a different trade-off between extero- and interoception by contrast with literature.

Seeing-In Pictures

The discussion of body schemas and pictures can also be couched in a broader question: what is the main difference with respect to the skills and habits that we bring to bear in pictorial perception compared to those we employ in the real world? Let's consider, for a moment, static images such as drawings, paintings, and photos. Such pictures are peculiar kinds of objects. I have argued elsewhere (Fingerhut, 2014, 2020a; Fingerhut and Heimann, 2017) that pictures (i) afford specific epistemic operations, that they are (ii) affective objects that can address us in powerful ways, and (iii) that via exposure and experience-based learning, we develop an artifact-specific perceptual manner of engagement with them. The latter aligns mostly with the topic of the present paper. To properly address our pictorial habits of perceiving, consider again the insight from enactive sense-making: cognizers must actively bring forth experiences. Enacting what we experience takes a different turn when we engage with pictures. The reason for this becomes obvious when we think about the sensorimotor patterns involved. Changing our position relative to the picture, for instance, does not allow us to see behind a depicted object. Pictures and depicted objects thus provide their own—and sometimes paradoxical—experiences of presence (Noë, 2012; Seth, 2019). Material pictures afford a different kind of exploration with respect to what is depicted (their content) and with respect to the properties of their surfaces (their configurational features). But most crucially, we experience a *surface-content relation* when we see a picture. In fact, it has been argued that perceiving pictures is constituted by engaging such a surface-content interaction; it relies on the cognitive operation of *seeing-in* that comes with the phenomenology of a *twofold* experience (Wollheim, 1980/2015; Hopkins, 2003; Lopes, 2003). To perceive something in a picture, we have to engage with its configurational and with its representational properties. Both jointly constitute the experience.

The intricacies of the philosophical debate regarding seeing-in are not relevant for the present paper as the point I would like to make is more general. It seems obvious here that perception of the surfaces of pictures and perception of what is depicted afford different sensorimotor operations. Yet it is the *interaction, parallel processing*, or *integration* of the two operations within the habit of picture perception, in particular, that must be better understood. This, strangely enough, is largely ignored in the cognitive sciences that use pictures as stimuli and even the field of neuroaesthetics (Fingerhut, 2018b).

Consider embodied simulation accounts that highlight motor responses as a necessary feature of our engagement with pictures such as paintings (Freedberg and Gallese, 2007). They focus on body postures, implied actions, and the facial expressions of depicted human figures on the one hand, and on premotor areas responding to perceived brushstrokes or cuts of the canvas on the other (Umiltà et al., 2012; Sbriscia-Fioretti et al., 2013). Nonetheless, they do not explore how both folds of our cognitive processing (i.e., of surface and content) interact in our engagement with a painting. That is, they do not show how the parallel motor processing of surface and content features determine our experiences in such cases.[21] I do not believe this is a minor point: if our perceptual habit of picture perception is defined by this double processing, then this is a necessary complication for any theory of pictorial engagement (Fingerhut, 2018a). This last point more generally attests to the need to study habits as a unit rather than as something constituted by disjointed processes. In order to understand picture perception, the intertwined processing of configuration and content afforded by those artifacts has to be taken into account.

It has been argued that film does not have a surface in the same way other pictures have and that therefore there is no seeing-in with respect to film (Cavell, 1979; Carroll, 1996). This can be illustrated by the central role of sensorimotor engagement with the surface of a handmade painting: moving toward a painting makes the brushstrokes more visible and might contribute to the central experience of the artwork (Currie, 2018). This does not occur in the same way with the surface on which film is shown, such as a projection screen in cinema. Nonetheless, there is good reason to extend the notion of seeing-in to moving images and the many screen-based digital media containing them. Also in film we interact with configurational features (edits, camera, and lens movements) and the evolving content simultaneously. As with representational static images, any account of our filmic habits would have to integrate this double engagement and explain how film actively guides our exploration through specific moving-image strategies (Fingerhut, 2020b).

Such a focus could constitute one way to complement the more generic features of our cognitive apparatus described in cognitive media theory. Yet it should come as no surprise that also other expansions of cognitive engagements through the medium of film have been explored in the literature. One example is empathy. It has been argued that film affords expansive empathic engagement by providing close-ups that, for instance, enable us to engage more intensely with the faces of depicted characters. This engagement facilitates a better understanding of people from what could be considered outgroups and to which we otherwise would not develop such an involvement.

[21] Embodied simulation accounts of pictures and pictorial artworks have been criticized for relying on inner representations as mediating such experiences and therefore not being properly embodied (Gallagher, 2011). I will not go into the details of this discussion here. But I believe that a more enactive understanding of the role of the motor system as involved in preparation for actions, as Gallagher suggests, might preserve some of the insights of the embodied simulation theory of the arts (Fingerhut, 2018b).

Smith (2012),BR171 discusses this within a 4E framework by referring to the aforementioned embodied simulation accounts (motor simulation of facial mimicry and observed actions). Embodied simulation functions as a mediator to enhance our engagement with characters that we would not have the same access to under normal conditions.

The kind of motor activity described by Smith is seen as having the domain-general function of facilitating empathy. Film thus expands some of the features (through close-ups of faces, gestures, etc.) we can pick up on as well as the class of organisms or objects (marginalized groups, aliens, robots, villains, etc.) to which we allot this kind of empathy. This is important in of itself. But what I want to add is that Smith's application of motor theories of empathy still relies on a bio- or socio-chauvinistic interpretation of neural activity. As I have argued above, such a view needs to be amended by a focus on the *neuromedial* elements that are part of the larger, structural way a movie recruits and engages the cognitive apparatus within our filmic habit. Motor activity is also modulated by filmic features such as camera movements and edits (Heimann et al., 2014, 2019) and therefore configurational features of the medium. We have to take into account how these have been incorporated into our ways of exploring a scene in film. This is what a *new cognitive media theory* should capitalize on. So in terms of the motor-empathy framework discussed in the preceding paragraphs, one could speak of "empathy with the medium"—one that not only includes the depicted persons *or* the stylistic means of film independent of each other, but centrally the integrated seeing-in habits related to moving images (i.e., the interplay of configuration and recognition in our engagement of film, see Fingerhut, 2020b). Any neural activity, and especially the neuromedial side of the larger artifactual habit, would have to be interpreted with respect to such normal conditions of film perception.

Digital and New Media

Pictures and moving images are intimately woven into recent digital revolutions. Concepts such as post-cinema or trans- and intermediality in storytelling capture only some of ways that images migrate or are processed therein. The presence of screen-based media is permanent both as portable devices and stable within our environment. Data from our interactions with such interfaces are fed back into what is presented on them (think of data-mining artificial intelligence in social media). The term 'new media' largely designates the field of social media, sometimes including the devices and gadgets used to engage with this particular media. But it also marks something that is akin to all media and fits the third notion of *expansiveness* from above: "by changing the conditions for the production of experience, new media destabilize existing patterns of biological, psychical, and collective life even as they furnish new facilities" (Hansen, 2010, p. 173). In this sense, old new media (the emergence of cave paintings, the printing press) might already reveal many things that can be applied to more recent new media as well (Manchovic, 2001), and could also help us understand our intensely digitally mediated environments.

In the expansive habits view I have proposed, new and digital media are interesting for many reasons. Such media create enhanced dynamics due to parallel available and transmedia ecologies that require an additional focus on the meta-habit of switching between multiple platforms, formats, and devices. However, I will focus on two central points only. First, digital media are not disembodied media. Their interfacing devices exploit existing embodied engagements by aiming to be more seamlessly integrated than other media have been to date. Second, media devices evolve in rapid reciprocal adjustments with users. Now, there are even media set-ups that employ real-time feedback loops and real-time adjustment to the organism. This relates to the growing domain of pervasive and ubiquitous computing in the background of our world (Lyytinen and Yoo, 2002) and to the algorithms and artificial intelligence (AI) used to predict our interests (as evidenced by various functions on Facebook, Twitter, Snapchat, TikTok, and so on). Such predictive activity emanating from the backend of media corresponds with the concept of *neuromediality* in an interesting way. Now, this concept denotes the neural contribution within a habit not only on the side of the organism but is also employed within the external medium itself. Today, media environments themselves operate under neural regimes.

Coming back to the first point I want to highlight. This involves embodied routines of interaction and the ways our bodily gestures (as well as those related to older media artifacts) became integrated into novel interfaces. Think of our use of touchscreens via gestures. A small but important point in this respect is that even such seemingly seamless devices do nonetheless require specific media skills (and related sensorimotor and body-schematic processes).

This has been demonstrated by developmental psychology and research into the so-called 'video deficit effect,' or the ability to transfer learned content from 2D to 3D to real-life-situations. Such transfer ability is relatively poor in infants (Anderson and Pempek, 2005). This means that media skills cannot be immediately applied in a domain-general way and as easily be transferred between media and outside the media context.

Recently, this kind of research has been extended to study what it means to grow up in new digital environments (Barr, 2019). Touchscreen devices appear to provide more interactive opportunities that should make transfer to 3D worlds outside the media context more immediate. Yet transfer deficits nonetheless remain also for touch screens. For example, children who learn to press buttons on a 2D touchscreen cannot use this skill with respect to 3D objects as immediately as one might expect (Zack et al., 2009). The overall point of such findings is that despite a general ability to transfer recognition and action skills between media, or between media and the real world, such transfers often come at a cost, such as additional cognitive load (Zack et al., 2013). While such a load seems to be neglectable and often remains unnoticed in adults, studies with infants provide some support for the claim that media habits require their own rules of engagement, even in media that seem to have adapted to the human motor-sensorium.

The second aforementioned aspect refers to the content and configuration of new digital media being adjusted in ever-shorter timescales (up to real time) to their users. A common example is learning software that adapts to the skillset of its user. Likewise, our choices determine the content portrayed to us in social media. Such responsive feedback is also at the heart of the concept of *enactive media* (Tikka, 2010a; Kaipainen et al., 2011). The structurally interesting features of such media is that they pick up on our actions and physiology and adjust their feedback accordingly. The authors describe one specific filmic media setting in which "technology is a part of a two-way feedback system with self-controlling recursive properties, and the role of an interface becomes implicit, perhaps even to the degree of being non-conscious" (Kaipainen et al., 2011, p. 433). The relevant cinema installation includes a montage machine unit that recombines elements from a database into cinematic composition based on psycho-physical data from the viewer (see also Tikka, 2010b). This makes the viewer the unconscious author of their media content.

Despite the focus on cinematic narrative, the discussion of enactive media has a more general relevance. For one, it makes explicit the possibility of new media systems to attune their user in real-time by in future also more systematically mining their physiological and neural data. For another, it simultaneously limits cognitive access to the interface of such adjustments. Much more could be said about whether enactive media introduce a new dynamicism from the artifact side, or whether they simply demonstrate more clearly how media always have entrained and transformed us. I included them in the present paper to demonstrate that a *new cognitive media theory* must not simply highlight media-specific abilities (artifactual habits beyond the generic cognitive abilities addressed by cognitive media theory). It must also address the dynamic reciprocal influences of organism and media environments, which both enactivism and RPP have made salient. Such dynamics might include a highly adaptive (and thus *neuromedially predictive*) element on the media artifact side as well. This element could change the character of media-related habits that already encompass artifact and organism (a I aimed to capture by the concept of *locational expansiveness*). Such neuromedial elements on the artifact side renders organism and media artifacts ever more intimately interwoven.

Architecture and Cities

Media have been treated as extensions of our bodies. McLuhan's media cases thus include buildings and cities, which are viewed as extensions of our metabolic system. But the built environment structures cognition and actions on a multitude of levels; it affects us continuously across all of our senses from vision to the vestibular, from touch to sound. We create our reality as we move through designed space. The impact of architecture and design remains a largely understudied field in philosophy and cognitive science. This is certainly true compared to study of language, but also compared to study of pictures and even computation and digitalization. Still, things have started to shift due in part to scholarly interest in the possible convergence of embodied cognition paradigms and architectural studies (Mallgrave, 2013; Pallasmaa et al., 2015; Robinson and Pallasmaa, 2017).

Currently, half of the world's population lives in densely populated urban areas. This portion is projected to rise above two thirds of the population by 2050. Recent studies have explored correlations between cities and mental health, noting that the risks for anxiety disorders and psychotic disorders such as schizophrenia might be significantly higher in cities (Gruebner et al., 2017; but see DeVylder et al., 2018). It thus seems pressing to study the impact of architecture and city planning, along with general urban cognitive ecosystems, on mental well-being, cognition, and experience. Some emerging fields, such as *neurourbanism*, do so (Adli et al., 2017; Fett et al., 2019). Essentially, the built environment is the ultimate designer environment for our embodied minds to fold into their cognizing and experiencing. This is because it is such a determining factor across a wide range of bodily actions.

It is worth briefly considering the constant and stabilizing influence of the built environment on our habits of engagement. Due to its continuous presence, we might overlook its impact. This would render architecture-related perceptual engagements a human constant that is no longer a visibly part of an artifactual habit. Still, there are some indications of how the built environment might have permeated our perception. One example is the Müller-Lyer illusion (which portrays two lines of equal length as different lengths to the human vision, thanks to fins at the end of the line protruding either outwards or inwards). The illusion appears to be universal. For instance, it is present in children who gain sight after congenital blindness (Gandhi et al., 2015). Yet the size of the effect is not universal. It has been smaller for Navajo native Americans who grew up in traditional roundhouses compared to those who grew up in new reservation architecture (Pedersen and Wheeler, 1983; Phillips, 2019). This has been related to a the 'carpentered world hypothesis'. The rationale is that we perceive lines with fins protruding outwards as being at the back of a room (or of something else in our carpentered worlds). They appear enlarged in our perception because the visual system compensates for them being seemingly further away.

Other studies have focused on the impact of navigation in cities on our cognitive system. In a seminal study on experience-driven neuroplasticity, taxi drivers in London showed greater gray matter volume in the mid-posterior hippocampi compared to bus drivers who do not have to exhibit the same navigational skills (Maguire et al., 2006). A more general exploration of the navigational capacities of 442,195 participants across 38 countries by the same lab found participants raised in cities had worse navigation skills than those raised in more rural areas. The effect was larger for cities that had a geometric grid layout compared to more organic and complex ones (Coutrot et al., 2020). The taxi driver data reflects a task-driven plasticity, while the city-rural comparison shows more generally how an environment recruits its organisms and then alters their cognitive capacities. The data therefore indicate that statistical immersion to an environment alters our embodied, cognitive habits and that organisms allocate neuronal processing resources (and undergo structural changes) according to the demands of their environments.

The co-dependency and reciprocal shaping of architectural and human embodiment also happens over smaller and dynamic timescales (Jelić et al., 2016). The stable presence of architectural

elements has a corollary effect (in keeping with the effects of precision weighing in PP discussed above) wherein small changes have rather big impacts. A central architectural element are entrances and doors that afford locomotive permeability. They have been, for instance, explored in EEG experiments that measure motor preparation in the perception of such apertures, which showed a highly fine-tuned sensitivity to this particular architectural element (i.e., whether a door is walk-through-able or not, see Djebbara et al., 2019, 2021). Such adjustments are part of our architectural, multisensory habit to perceive architectural affordances. Within such sensorimotor engagements, we can understand how our experience of the built environment unfolds. Here, we pick up on a multitude of design decisions and architectural features in a dynamic way.[22]

After the initial interest from McLuhan (1964, 1988), buildings and cities did not become a central concern for media theory (but see Kittler and Griffin, 1996). As multisensory and mixed media environments, cities and architecture have re-entered the media theory landscape only recently (McQuire, 2008). Part of the reason for their return is the rise of ubiquitous and pervasive computing in smart cities. Artificial intelligence, the *Internet of Things*, and large-scale data analytics are now employed to predict and influence behavior. In this context, "architecture provides a fixed form for the flows engineered by pervasive computing" (McCullough, 2007, p. 395). Social media for city experiences (Molinillo et al., 2019) and sensory feedback loops in buildings might themselves become a central part of what we consider architecture in the future as they latch onto our already artifactual habits of engagement.

This section has described how artifactual habits relating to urban and architectural design entrain our perceptual engagement and determine cognitive capacities. The built environment presents us with experiential models in ways that are comparable to other media. Design decisions and urban planning provide different models for how we may live together. They influence urban dwellers in terms of their social behavior or explorations of their environment. By focusing on how design decision nudge us in cities and buildings (even without their 'smart' extensions), those cities could be described as media. They process, store, and transmit information, yet over longer timescales compared to other media. At the same time, they are projections of the kind of social being that a certain culture aims to produce and promote (for some critical implications of this, see Crippen and Klement, 2020). As such, architecture and the built environment are models of who we are (have been and will be) as a society.

CONCLUSION AND OUTLOOK

Media environments and technologies evolve with our embodied brain-body nexus in reciprocal co-adaptations. In this, they constantly reconfigure and transform how we engage and experience. I have aimed to capture some of these dynamics by highlighting the expansive artifactual habits we entertain (because we live in a built environment, among a plethora of pictures, and now are immersed in new digital media that respond dynamically to us). I have mainly discussed media artifacts from pictorial domains at the omission of other elements or mixed media environments (such as sound and spoken language, texts, and how those are interwoven and interact with pictures) because I see images as underrepresented in the discussion of the relation of culture and mind. But I also aimed at a more general point: media in all their ramifications should occupy a central place within the still-maturing field of situated cognition.

I have therefore focused on a rather general concept in the philosophy of mind, namely habits (Caruana and Testa, 2020). With this, I sought to capture the basic insight into the relational nature of our mind propagated by 4E and enculturation theories alike: our mind is crucially determined by the embodied actions afforded by our socio-techno-cultural environments. As I introduced them, habits are critical qualifiers of the range of such actions within a specific ecology. In media ecologies, we are expert perceivers without knowing it. The way we explore the contents of different media is couched in habits that are partially constituted by the structural features of the media artifacts themselves. They are not rigid mechanical routines. Instead, habits are flexible ways of world-making.

I have only briefly tapped into the rich and evolving field of media studies by highlighting some general claims regarding media archeology and ecology. More specifically, I have addressed the way cognitive media theory captures our media engagement. Although this media theory has recently started to include ideas from situated cognition, I suggest that there are limitations to this account. In comparison, the pluralistic and dynamic view of artifactual habits (along with the interlocking of media and neuro-cognitive architecture) in my enactive account of media constitutes a larger shift in thinking. This shift might warrant the label of *new cognitive media theory*. Regardless, it entails acknowledgment of the plurality of habits and related bodily engagements (I discussed the filmic body schema we entertain when engaging with the pervasive artifact of moving images, as well as the capacity of seeing-in that pertains to all pictorial domains). It further offers an ensuing understanding of how our perceptual, emotional, and aesthetic engagement unfolds *within* such habits based on new insights into our cognitive apparatus.

No survey of situated or 4E accounts can be exhaustive. The field has evolved so rapidly that one is liable to miss out on developments even for subdomains like media engagement, which – unduly to my mind – are treated only at its periphery (I am, for instance, well-aware that I largely ignored phenomenological and post-phenomenological thinking regarding media). I aimed to capture some central junctures to the artifactual habits account of media I propose. Thus, I aimed to re-territorialize extended mind claims to sociocultural media-ecologies while retaining some of their focus on mixed-media coalitions *within* habits. I did not focus on the ontological claims

[22]This can be illustrated by the impact of sound within the built environment. For example, sonic feedback from our own movements (manipulated to low vs. high pitch) can influence how large or heavy we experience our body to be (Tajadura-Jiménez et al., 2015). This demonstrates our capacity for multisensory, fine-tuned adjustment based on normal conditions within an architectural habit.

related to this. Instead, I proposed an enactive understanding of how cultural artifacts have become integrated into our cognitive routines. As central element, they do so by bringing forth experiences in domains that sustain their own rules and values. I argued that radical predictive processing (RPP) could provide an accompanying explanation of how the nervous system facilitates organism-artifact coalitions and how we attune to design environments on multiple levels.

Our ability to engage with a plurality of designed media models captures something central and defining in human cognizing and experiencing. Once we understand the expansive artifactual habits that bring forth novel meanings and values, we can understand how our mind is mediated and becomes re-mediated at every moment of being engaged with such models. RPP served to situate the more local neuronal contribution within this larger picture; it elucidates a possible role of the brain in folding designed, media environments into our embodied engagements. Further, the concept of *neuromediality* captures some of this. It brings into focus the exapted functions certain neuronal processes might take on in different media ecologies. As such, neuromedial processes are part of the normal conditions of any media engagement. In recent digital media developments, neuromedial processes could even be ascribed to media themselves (as we saw with respect to the real-time dynamics of adapting and predicting their users).

This paper aimed to contribute to a broader understanding of enculturation in situated cognition by focusing on how we actively bring forth experiential models of the world that become salient through and within media. It did not address what could be considered our aesthetic relations to such cultural artifacts. Media and cultural artifacts actively invite our exploration of the world. They also invite evaluation of their ways of worldmaking. Aesthetic and emotional appreciation might be a central way to track the bundles of perceptual, cognitive, and other effects presented to us by cultural artifacts (I explore such relations elsewhere, see Fingerhut, 2018b; Fingerhut and Prinz, 2020, forthcoming). Aesthetic evaluations of specific media outputs relate to normative claims. This poses a threat to a more comprehensive convergence between the humanities element in media studies and naturalistic explanations in the 4E cognitive sciences (Nannicelli, 2019). Future research will have to address this. One promising way could be to explore what the present paper has established as the more general value-generating enactive view of habits and the affective dimension of the respective media models this entails.

DATA AVAILABILITY STATEMENT

The original contributions generated for this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

Work on this manuscript was made possible by support from the Einstein Foundation Berlin and the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 870827: ARTIS. I also acknowledge support by the German Research Foundation (DFG) and the Open Access Publication Fund of Humboldt-Universität zu Berlin.

ACKNOWLEDGMENTS

Special thanks to Matthew Crippen for helpful comments and a thorough reading of the manuscript. Thanks to Inês Hipólito and Riccardo Manzotti for helpful feedback on earlier versions and to Corinna Kühnapfel for help with the references. Special thanks also to the two referees who prompted me to include additional literature and steered the manuscript toward more clarity.

REFERENCES

Adams, F. (2003). The informational turn in philosophy. *Minds Mach.* 13, 471–501. doi: 10.1023/A:1026244616112

Adams, F., and Aizawa, K. (2001). The bounds of cognition. *Philos. Psychol.* 14, 43–64. doi: 10.1080/09515080120033571

Adams, R. A., Shipp, S., and Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Struct. Funct.* 218, 611–643. doi: 10.1007/s00429-012-0475-5

Adli, M., Berger, M., Brakemeier, E.-L., Engel, L., Fingerhut, J., Gomez-Carrillo, A., et al. (2017). Neurourbanism: towards a new discipline. *Lancet Psychiatry* 4, 183–185. doi: 10.1016/S2215-0366(16)30371-6

Anderson, D. R., and Pempek, T. A. (2005). Television and very young children. *Am. Behav. Sci.* 48, 505–522. doi: 10.1177/0002764204271506

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–313. doi: 10.1017/S0140525X10000853

Anderson, M. L., and Chemero, A. (2018). "The world well gained: on the epistemic implications of ecological information," in *Andy Clark and His Critics*, eds M. Colombo, E. Irvine, and M. Stapleton (Oxford: Oxford University Press), 161–173.

Barr, R. (2019). Growing up in the digital age: early learning and family media ecology. *Curr. Dir. Psychol. Sci.* 28, 341–346. doi: 10.1177/0963721419838245

Bassolino, M., Serino, A., Ubaldi, S., and Làdavas, E. (2010). Everyday use of the computer mouse extends peripersonal space representation. *Neuropsychologia* 48, 803–811. doi: 10.1016/j.neuropsychologia.2009.11.009

Berliner, T., and Cohen, D. J. (2011). The illusion of continuity: active perception and the classical editing system. *J. Film Video* 63, 44–63. doi: 10.5406/jfilmvideo.63.1.0044

Bertolotti, T., and Magnani, L. (2017). Theoretical considerations on cognitive niche construction. *Synthese* 194, 4757–4779. doi: 10.1007/s11229-016-1165-2

Bickenbach, M. (2011). Blindness or insight? Kittler on culture. *Thesis Eleven* 107, 39–46. doi: 10.1177/0725513611418038

Biederman, I., and Vessel, E. (2006). Perceptual pleasure and the brain: a novel theory explains why the brain craves information and seeks it through the senses. *Am. Sci.* 94, 247–253.

Bordwell, D. (2013). "The viewer's share: models of mind in explaining film," in *Psychocinematics: Exploring the Cognition at the Movies*, ed. A. P. Shimamura (New York, NY: Oxford University Press), 29–52. doi: 10.1093/acprof:oso/9780199862139.003.0002

Bourdieu, P. (1977). *Outline of a Theory of Practice*. Cambridge, MA: Cambridge University Press.

Brumm, A., Oktaviana, A. A., Burhan, B., Hakim, B., Lebe, R., Zhao, J. X., et al. (2021). Oldest cave art found in Sulawesi. *Sci. Adv.* 7:eabd4648. doi: 10.1126/sciadv.abd4648

Carroll, N. (1985). The power of movies. *Daedalus* 114, 79–103.

Carroll, N. (1996). *Theorizing the Moving Image*. Cambridge, MA: Cambridge University Press.

Carroll, N., and Seeley, W. P. (2013). "Cognitivism, psychology, and neuroscience: movies as attentional engines," in *Psychocinematics*, ed. A. P. Shimamura (New York, NY: Oxford University Press), 53–75. doi: 10.1093/acprof:oso/9780199862139.003.0003

Caruana, F., and Testa, I. (eds) (2020). *Habits: Pragmatist Approaches from Cognitive Science, Neuroscience, and Social Theory*. Cambridge, MA: Cambridge University Press.

Cavell, S. (1979). *The World Viewed Reflections on the Ontology of Film (Enlarged Edition)*. Cambridge, MA: Harvard University Press.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Christensen, W., Sutton, J., and McIlwain, D. J. F. (2016). Cognition in skilled action: meshed control and the varieties of skill experience. *Mind Lang.* 31, 37–66. doi: 10.1111/mila.12094

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. (1998). Being there: putting philosopher, researcher and student together again (author's response. *Metascience* 7, 95–104. doi: 10.1007/bf02913277

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Clark, A. (2015a). Radical predictive processing. *South. J. Philos.* 53, 3–27. doi: 10.1111/sjp.12120

Clark, A. (2015b). Predicting peace the end of the representation wars – a reply to Michael Madary. *Open MIND* 7, 1–7. doi: 10.15502/9783958570979

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. New York, NY: Oxford University Press.

Clark, A. (2018). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenom. Cogn. Sci.* 17, 521–534. doi: 10.1007/s11097-017-9525-z

Clark, A. (2019). "Replies to critics: search of the embodied, extended, enactive, predictive (EEE-P) mind," in *Andy Clark and His Critics*, eds M. Colombo, E. Irvine, and M. Stapleton (Oxford: Oxford University Press), 266–302.

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 7–19. doi: 10.1093/analys/58.1.7

Clowes, R. W. (2019). Immaterial engagement: human agency and the cognitive ecology of the internet. *Phenomenol. Cogn. Sci.* 18, 259–279. doi: 10.1007/s11097-018-9560-4

Colombetti, G. (2017). Enactive affectivity, extended. *Topoi* 36, 445–455. doi: 10.1007/s11245-015-9335-2

Constant, A., Clark, A., Kirchhoff, M., and Friston, K. (2020a). Extended active inference: constructing predictive cognition beyond skulls. *Mind Lang.* 12, 1–22. doi: 10.1111/mila.12330

Constant, A., Tschantz, A., Millidge, B., Criado-boado, F., and Clark, A. (2020b). The acquisition of culturally patterned attention styles under active inference. *PsyArXiv* [Preprint] doi: 10.31234/osf.io/rchaf

Coutrot, A., Manley, E., Yesiltepe, D., Dalton, R. C., Wiener, J. M., Hölscher, C., et al. (2020). Cities have a negative impact on navigation ability: evidence from 38 countries. *bioRxiv* [Preprint] doi: 10.1101/2020.01.23.917211

Crippen, M. (2020). Enactive pragmatism and ecological psychology. *Front. Psychol.* 11:538644. doi: 10.3389/fpsyg.2020.538644

Crippen, M., and Klement, V. (2020). Architectural values, political affordances and selective permeability. *Open Philos.* 3, 462–477. doi: 10.1515/opphil-2020-0112

Crogan, P., and Kinsley, S. (2012). Paying attention: towards a critique of the attention economy. *Cult. Mach.* 13, 1–29.

Cuffari, E. C., Di Paolo, E., and De Jaegher, H. (2015). From participatory sense-making to language: there and back again. *Phenom. Cogn. Sci.* 14, 1089–1125. doi: 10.1007/s11097-014-9404-9

Currie, G. (2018). "Pictures and their surfaces," in *The Pleasure of Pictures: Pictorial Experience and Aesthetic Appreciation*, eds J. Pelletier and A. Voltolini (New York, NY: Routledge), 330–359. doi: 10.4324/9781315112640-14

De Jaegher, H., and Di Paolo, E. (2007). Participatory sense-making: an enactive approach to social cognition. *Phenom. Cogn. Sci.* 6, 485–507. doi: 10.1007/s11097-007-9076-9

Dennett, D. C. (1993). The message is: there is no medium. *Philos. Phenomenol. Res.* 53, 919–931. doi: 10.2307/2108264

Dennett, D. C. (2001). Are we explaining consciousness yet? *Cognition* 79, 221–237. doi: 10.1016/s0010-0277(00)00130-x

D'Errico, F., and Colagè, I. (2018). Cultural exaptation and cultural neural reuse: a mechanism for the emergence of modern culture and behavior. *Biol. Theory* 13, 213–227. doi: 10.1007/s13752-018-0306-x

DeVylder, J. E., Kelleher, I., Lalane, M., Oh, H., Link, B. G., and Koyanagi, A. (2018). Association of urbanicity with psychosis in low- and middle-income countries. *JAMA Psychiatry* 75:678. doi: 10.1001/jamapsychiatry.2018.0577

Dewey, J. (1983). "Human nature and conduct," in *The Middle Works of John Dewey*, ed. A. J. Boydston (Carbondale, IL: Southern Illinois University Press), 226.

Di Paolo, E. (2009). Extended life. *Topoi* 28, 9–21. doi: 10.1007/s11245-008-9042-3

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenom. Cog. Sci.* 4, 429–452. doi: 10.1007/s11097-005-9002-y

Di Paolo, E., and Thompson, E. (2014). "The enactive approach," in *The Routledge Handbook of Embodied Cognition*, ed. L. Shapiro (New York, NY: Taylor & Francis Group), 68–78.

Di Paolo, E. A., Cuffari, E. C., and De Jaegher, H. (2018). *Linguistic Bodies: The Continuity Between Life and Language*. Cambridge, MA: MIT Press.

Djebbara, Z., Fich, L. B., and Gramann, K. (2021). The brain dynamics of architectural affordances during transition. *Sci. Rep.* 11:2796. doi: 10.1038/s41598-021-82504-w

Djebbara, Z., Fich, L. B., Petrini, L., and Gramann, K. (2019). Sensorimotor brain dynamics reflect architectural affordances. *Proc. Natl. Acad. Sci. U.S.A.* 116, 14769–14778. doi: 10.1073/pnas.1900648116

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press.

Dotov, D. G., Nie, L., and Chemero, A. (2010). A demonstration of the transition from ready-to-hand to unready-to-hand. *PLoS One* 5:e9433. doi: 10.1371/journal.pone.0009433

Dretske, F. I. (1981). *Knowledge and the Flow of Information*. Oxford: Basil Blackwell.

Durt, C., Tewes, C., and Fuchs, T. (2017). *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*. Cambridge, MA: MIT Press.

Fabry, R. E. (2018). Betwixt and between: the enculturated predictive processing approach to cognition. *Synthese* 195, 2483–2518. doi: 10.1007/s11229-017-1334-y

Feiten, T. E. (2020). Mind after uexküll: a foray into the worlds of ecological psychologists and enactivists. *Front. Psychol.* 11:480. doi: 10.3389/fpsyg.2020.00480

Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215

Fett, A.-K. J., Lemmers-Jansen, I. L., and Krabbendam, L. (2019). Psychosis and urbanicity: a review of the recent literature from epidemiology to neurourbanism. *Curr. Opin. Psychiatry* 32, 232–241. doi: 10.1097/YCO.0000000000000486

Fingerhut, J. (2012). "The body and the experience of presence," in *Feelings of Being Alive*, eds J. Fingerhut and S. Marienberg (Berlin: De Gruyter), 167–199. doi: 10.1515/9783110246599.167

Fingerhut, J. (2014). "Extended imagery, extended access, or something else? Pictures and the extended mind hypothesis," in *Bildakt at the Warburg Institute*, eds S. Marienberg and J. Trabant (Berlin: De Gruyter), 33–50. doi: 10.1515/9783110364804.33

Fingerhut, J. (2018a). Embodied seeing-in, empathy, and expansionism. *Projections* 12, 28–38. doi: 10.3167/proj.2018.120205

Fingerhut, J. (2018b). Enactive aesthetics and neuroaesthetics. *Phenomenol. Mind* 14, 80–97. doi: 10.13128/Phe_Mi-23627

Fingerhut, J. (2020a). "Habits and the enculturated mind: pervasive artifacts, predictive processing, and expansive habits," in *Habits: Pragmatist Approaches from Cognitive Neuroscience to Social Science*, eds F. Caruana and I. Testa (Cambridge, MA: Cambridge University Press), 352–375.

Fingerhut, J. (2020b). Twofoldness in moving images. On the philosophy and neuroscience of filmic experiences. *Projections* 14, 1–20. doi: 10.3167/proj.2020.140302

Fingerhut, J., and Heimann, K. (2017). "Movies and the mind: on our filmic body," in *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*, eds C. Durt, T. Fuchs, and C. Tewes (Cambridge, MA: MIT Press), 353–377.

Fingerhut, J., and Prinz, J. J. (2018). Wonder, appreciation, and the value of art. *Prog. Brain Res.* 237, 107–128. doi: 10.1016/bs.pbr.2018.03.004

Fingerhut, J., and Prinz, J. J. (2020). Aesthetic emotions reconsidered. *Monist* 103, 223–239. doi: 10.1093/monist/onz037

Fingerhut, J., and Prinz, J. J. (forthcoming). *Enactive Aesthetic Emotions*.

Fox, J., and McEwan, B. (2017). Distinguishing technologies for social interaction: the perceived social affordances of communication channels scale. *Commun. Monogr.* 84, 298–318. doi: 10.1080/03637751.2017.1332418

Freedberg, D., and Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends Cogn. Sci.* 11, 197–203. doi: 10.1016/j.tics.2007.02.003

Fridland, E. (2017). Skill and motor control: intelligence all the way down. *Philos. Stud.* 174, 1539–1560. doi: 10.1007/s11098-016-0771-7

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. B* 364, 1211–1221. doi: 10.1098/rstb.2008.0300

Friston, K., Thornton, C., and Clark, A. (2012). Free-energy minimization and the dark-room problem. *Front. Psychol.* 3:130. doi: 10.3389/fpsyg.2012.00130

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybere.* 102, 227–260. doi: 10.1007/s00422-010-0364-z

Fuchs, T. (2011). The brain–A mediating organ. *J. Conscious. Stud.* 18, 196–221.

Gallagher, S. (2011). "Aesthetics and kinaesthetics," in *Sehen und Handeln*, eds S. Marienberg and J. Trabant (Berlin: De Gruyter), 99–113. doi: 10.1524/9783050062389.99

Gallagher, S., and Allen, M. (2018). Active inference, enactivism and the hermeneutics of social cognition. *Synthese* 195, 2627–2648. doi: 10.1007/s11229-016-1269

Gallagher, S., and Varga, S. (2020). Meshed architecture of performance as a model of situated cognition. *Front. Psychol.* 11:2140. doi: 10.3389/fpsyg.2020.02140

Gandhi, T., Kalia, A., Ganesh, S., and Sinha, P. (2015). Immediate susceptibility to visual illusions after sight onset. *Curr. Biol.* 25, R358–R359. doi: 10.1016/j.cub.2015.03.005

Gane, N. (2005). Radical post-humanism: Friedrich Kittler and the primacy of technology. *Theory Cult. Soc.* 22, 25–41. doi: 10.1177/0263276405053718

Gane, N., and Sale, S. (2007). Interview with Friedrich Kittler and Mark Hansen. *Theory Cult. Soc.* 24, 323–329. doi: 10.1177/0263276407086401

Gendron, M., Mesquita, B., and Barrett, L. F. (2020). "The brain as a cultural artifact: concepts, actions, and experiences within the human affective niche," in *Culture, Mind, and Brain*, eds L. J. Kirmayer, C. M. Worthman, S. Kitayama, R. Lemelson, and C. Cummings (Cambridge, MA: Cambridge University Press), 188–222. doi: 10.1017/9781108695374.010

Gruebner, O., Rapp, M., Adli, M., Kluge, U., Galea, S., and Heinz, A. (2017). Cities and mental health. *Dtsch. Ärztebl. Int.* 114, 121–127. doi: 10.3238/arztebl.2017.0121

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum.

Halpin, H., Clark, A., and Wheeler, M. (2010). "Towards a philosophy of the web: representation, enaction, collective intelligence," in *Proceedings of the 2nd Web Science Conference*, eds H. Halpin and A. Monnin (Raleigh, NC), 1–5. doi: 10.1002/9781118700143.ch2

Hansen, M. B. (2010). "New media," in *Critical Terms for Media Studies*, eds W. J. T. Mitchell and M. B. N. Hansen (Chicago, IL: University of Chicago Press), 172–185.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. J. (2008). Neurocinematics: the neuroscience of film. *Projections* 2, 1–26. doi: 10.3167/proj.2008.020102

Heimann, K., Uithol, S., Calbi, M., Umiltà, M. A. M. A., Guerra, M., Fingerhut, J., et al. (2019). Embodying the camera: an EEG study on the effect of camera movements on film spectators'; sensorimotor cortex activation. *PLoS One* 14:e0211026. doi: 10.1371/journal.pone.0211026

Heimann, K., Umiltà, M. A., Guerra, M., and Gallese, V. (2014). Moving mirrors: a high-density EEG study investigating the effect of camera movements on motor cortex activation during action observation. *J. Cogn. Neurosci.* 26, 2087–2101. doi: 10.1162/jocn_a_00602

Heimann, K. S., Uithol, S., Calbi, M., Umiltà, M. A., Guerra, M., and Gallese, V. (2017). Cuts in action: a high density EEG study investigating the neural correlates of different editing techniques in film. *Cogn. Sci.* 41, 1555–1588. doi: 10.1111/cogs.12439

Herbec, A., Kauppi, J.-P., Jola, C., Tohka, J., and Pollick, F. E. (2015). Differences in fMRI intersubject correlation while viewing unedited and edited videos of dance performance. *Cortex* 71, 341–348. doi: 10.1016/j.cortex.2015.06.026

Heyes, C. (2012). New thinking: the evolution of human cognition. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 2091–2096. doi: 10.1098/rstb.2012.0111

Heyes, C. (2020). Culture. *Curr. Biol.* 30, 1246–1250. doi: 10.1016/j.cub.2020.08.086

Hipólito, I., Baltieri, M., Friston, J. K., and Ramstead, M. J. (2020). *Embodied Skillful Performance: Where the Action is*. Available online at: http://philsci-archive.pitt.edu/17280/ (Accessed December 13, 2020)

Hopkins, R. (2003). Pictures, phenomenology and cognitive science. *Monist* 86, 653–675. doi: 10.5840/monist200386434

Hutchins, E. (2008). The role of cultural practices in the emergence of modern human intelligence. *Philos. Trans. R. Soc. B* 363, 2011–2019. doi: 10.1098/rstb.2008.0003

Hutchins, E. (2011). Enculturating the supersized mind. *Philos. Stud.* 152, 437–446. doi: 10.1007/s11098-010-9599-8

Hutto, D. D. (2005). Knowing what? Radical versus conservative enactivism. *Phenomenol. Cogn. Sci.* 4, 389–405. doi: 10.1007/s11097-005-9001-z

Hutto, D. D. (2018). Getting into predictive processing's great guessing game: bootstrap heaven or hell? *Synthese* 195, 2445–2458. doi: 10.1007/s11229-017-1385-0

Hutto, D. D., Gallagher, S., Ilundain-Agurruza, J., and Hipólito, I. (2020). "Culture in mind—an enactivist account: not cognitive penetration but cultural permeation," in *Culture, Mind, and Brain: Emerging Concepts, Models, Applications*, eds L. J. Kirmayer, S. Kitayama, C. M. Worthman, R. Lemelson, and C. Cummings (Cambridge, MA: Cambridge University Press), 163–187.

Hutto, D. D., and Myin, E. (2012). "Radicalizing enactivism: basic minds without content," in *Radicalizing Enactivism: Basic Minds without Content*, (MIT Press). doi: 10.1093/pq/pqt032

Ildirar, S., and Ewing, L. (2018). Revisiting the Kuleshov effect with first-time viewers. *Projections* 12, 19–38. doi: 10.3167/proj.2018.120103

Ildirar, S., and Schwan, S. (2015). First-time viewers' comprehension of films: bridging shot transitions. *Br. J. Psychol.* 106, 133–151. doi: 10.1111/bjop.12069

Jelić, A., Tieri, G., De Matteis, F., Babiloni, F., and Vecchiato, G. (2016). The enactive approach to architectural experience: a neurophysiological perspective on embodiment, motivation, and affordances. *Front. Psychol.* 7:481. doi: 10.3389/fpsyg.2016.00481

Jones, C. (2010). "Senses," in *Critical Terms for Media Studies*, eds W. J. T. Mitchell and M. B. N. Hansen (Chicago, IL: University of Chicago Press), 88–100.

Kaipainen, M., Ravaja, N., Tikka, P., Vuori, R., Pugliese, R., Rapino, M., et al. (2011). Enactive systems and enactive media: embodied human-machine coupling beyond interfaces. *Leonardo* 44, 433–438. doi: 10.1162/LEON_a_00244

Katkoff, M. (2019). *2019 Predictions: #1 only one way to climb to the top of the puzzle games category. Deconstructor of Fun*. Available online at: https://www.deconstructoroffun.com/blog/2019/1/14/2019-predictions-casual-games (accessed October 30, 2020)

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., and Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *J. R. Soc. Interface* 15:20170792. doi: 10.1098/rsif.2017.0792

Kirchhoff, M. D. (2012). Extended cognition and fixed properties: steps to a third-wave version of extended cognition. *Phenomenol. Cogn. Sci.* 11, 287–308. doi: 10.1007/s11097-011-9237-8

Kirchhoff, M. D. (2015). Extended Cognition & the causal-constitutive fallacy: in search for a diachronic and dynamical conception of constitution. *Philos. Phenomenol. Res.* 90, 320–360. doi: 10.1111/phpr.12039

Kirchhoff, M. D., and Kiverstein, J. (2019). *Extended Consciousness and Predictive Processing: A Third Wave View*. London: Routledge.

Kirchhoff, M. D., and Kiverstein, J. (2020). Attuning to the world: the diachronic constitution of the extended conscious ind. *Front. Psychol.* 11:1966. doi: 10.3389/fpsyg.2020.01966

Kirmayer, L. J., Worthman, C. M., and Kitayama, S. (2020). "Introduction," in *Culture, Mind, and Brain: Emerging Concepts, Models, Applications*, eds L. J. Kirmayer, S. Kitayama, C. M. Worthman, R. Lemelson, and C. Cummings (Cambridge, MA: Cambridge University Press), 1–50. doi: 10.1017/9781108695374.002

Kirsh, D., and Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cogn. Sci.* 18, 513–549. doi: 10.1207/s15516709cog1804_1

Kittler, F. A. (1999). *Gramophone, Film, Typewriter*. Stanford, CA: Stanford University Press.

Kittler, F. A., and Griffin, M. (1996). The city is a medium. *New Lit. Hist.* 27, 717–729.

Kukkonen, K. (2014). Presence and prediction. The embodied reader's cascades of cognition. *Style* 48, 367–384. doi: 10.5325/style.48.3.367

Kukkonen, K. (2020). *Probability Designs: Literature and Predictive Processing*. Oxford: Oxford University Press, doi: 10.1093/oso/9780190050955.001.0001

Làdavas, E. (2002). Functional and dynamic properties of visual peripersonal space. *Trends Cogn. Sci.* 6, 17–22. doi: 10.1016/S1364-6613(00)01814-3

Logan, R. K. (2013). McLuhan extended and the extended mind thesis (EMT). *Avant* 4, 45–58. doi: 10.12849/40202013.0709.0003

Lopes, D. M. (2003). Pictures and the representational mind. *Monist* 86, 632–652. doi: 10.5840/monist200386432

Loschky, L. C., Larson, A. M., Magliano, J. P., and Smith, T. J. (2015). What would jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLoS One* 10:e0142474. doi: 10.1371/journal.pone.0142474

Lum, C. M. K. (2014). "Media ecology," in *The Handbook of Media and Mass Communication Theory*, eds R. S. Fortner and F. P. Mark (Hoboken, NJ: John Wiley & Sons, Ltd), 137–153. doi: 10.1002/9781118591178.ch8

Lupyan, G., and Clark, A. (2015). Words and the world: predictive coding and the language-perception-cognition interface. *Curr. Dir, Psychol. Sci.* 24, 279–284. doi: 10.1177/0963721415570732

Lyytinen, K., and Yoo, Y. (2002). Ubiquitous computing. *Commun. ACM* 45, 62–65. doi: 10.1145/585597.585616

Maffongelli, L., Bartoli, E., Sammler, D., Koelsch, S., Campus, C., Olivier, E., et al. (2015). Distinct brain signatures of content and structure violation during action observation. *Neuropsychologia* 75, 30–39. doi:10.1016/j.neuropsychologia.2015.05.020

Magliano, J. P., and Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cogn. Sci.* 35, 1489–1517. doi: 10.1111/j.1551-6709.2011.01202.x

Maguire, E. A., Woollett, K., and Spiers, H. J. (2006). London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. *Hippocampus* 16, 1091–1101. doi: 10.1002/hipo.20233

Mallgrave, H. F. (2013). *Architecture and Embodiment: The Implications of the New Sciences and Humanities for Design*. London: Routledge.

Manchovic, L. (2001). *The Language of the New Media*. Cambridge, MA: MIT Press.

McCullough, M. (2007). New media urbanism: grounding ambient information technology. *Environ. Plann. B: Plann. Des.* 34, 383–395. doi: 10.1068/b32038

McLuhan, M. (1962). *The Gutenberg Galaxy: "When Change Becomes the Fate of Man"*. Toronto: University of Toronto Press.

McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. New York, NY: McGraw-Hill.

McLuhan, M. (1988). *Laws of Media: The New Science*. Toronto: University of Toronto Press.

McQuire, S. (2008). *The Media City: Media, Architecture and Urban Space*. London: SAGE Publications Ltd, doi: 10.4135/9781446269572

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Basingstoke: Palgrave Macmillan.

Menary, R. (2015). Mathematical cognition – a case of enculturation. *Open MIND* 25, 12–18. doi: 10.15502/9783958570818

Menary, R. (2018). "Cognitive integration: how culture transforms us and extends our cognitive capabilities," in *The Oxford Handbook of 4E Cognition*, eds S. Gallagher, A. Newen, and L. De Bruin (Oxford: Oxford University Press), 874–890. doi: 10.1093/oxfordhb/9780198735410.013.47

Miyahara, K., Ransom, T. G., and Gallagher, S. (2020). "What the situation affords. habit and heedful interrelations in skilled performance," in *Habits: Pragmatist Approaches from Cognitive Neuroscience to Social Science*, eds F. Caruana and I. Testa (Cambridge, MA: Cambridge University Press), 120–136. doi: 10.1017/9781108682312.006

Molinillo, S., Anaya-Sánchez, R., Morrison, A. M., and Coca-Stefaniak, J. A. (2019). Smart city communication via social media: analysing residents' and visitors' engagement. *Cities* 94, 247–255. doi: 10.1016/j.cities.2019.06.003

Nannicelli, T. (2019). Aesthetics and the limits of the extended mind. *Br. J. Aeshtetic* 59, 81–94. doi: 10.1093/aesthj/ayy048

Nannicelli, T., and Taberham, P. (2014). *Cognitive Media Theory*. New York, NY: Routledge, doi: 10.4324/9780203098226

Noë, A. (2009). Conscious reference. *Philos. Q.* 59, 470–482. doi: 10.1111/j.1467-9213.2009.630.x

Noë, A. (2012). *Varieties of Presence*. Cambridge, MA: Harvard University Press.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973. doi: 10.1017/S0140525X01000115

Pallasmaa, J., Mallgrave, H. F., and Arbib, M. (2015). Architecture and neuroscience. *Archit. Res. Q.* 19, 361–367. doi: 10.1017/S1359135515000627

Parola, M., Johnson, S., and West, R. (2016). Turning presence inside-out: MetaNarratives. *J. Electron. Imaging* 4, 1–9. doi: 10.2352/ISSN.2470-1173.2016.4.ERVR-418

Parr, T., and Friston, K. J. (2017). Working memory, attention, and salience in active inference. *Sci. Rep.* 7:14678. doi: 10.1038/s41598-017-15249-0

Pearlman, K. (2017). Editing and cognition beyond continuity. *Projections* 11, 67–86. doi: 10.3167/proj.2017.110205

Pedersen, D. M., and Wheeler, J. (1983). The Müller-Lyer Illusion among Navajos. *J. Soc. Psychol.* 121, 3–6. doi: 10.1080/00224545.1983.9924459

Perez-Marcos, D., Bieler-Aeschlimann, M., and Serino, A. (2018). Virtual reality as a vehicle to empower motor-cognitive neurorehabilitation. *Front. Psychol.* 9:2120. doi: 10.3389/fpsyg.2018.02120

Phillips, W. L. (2019). "Cross-cultural differences in visual perception of color, illusions, depth, and pictures," in *Cross-Cultural Psychology*, ed. K. D. Keith (Hoboken, NJ: John Wiley & Sons, Ltd), 287–308. doi: 10.1002/9781119519348.ch13

Pisters, P. (2017). *The Neuro-Image: A Deleuzian Film-Philosophy of Digital Screen Culture*. Stanford: Stanford University Press.

Prinz, J. J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Ramstead, M. J., Veissière, S., and Kirmayer, L. (2016). Cultural affordances: scaffolding local worlds through shared intentionality and regimes of attention. *Front. Psychol.* 7:1090. doi: 10.3389/fpsyg.2016.01090

Rietveld, E., and Kiverstein, J. (2014). A rich landscape of affordances. *Ecol. Psychol.* 26, 325–352. doi: 10.1080/10407413.2014.958035

Robinson, S., and Pallasmaa, J. (2017). *Mind in Architecture: Neuroscience, Embodiment, and the Future of Design. Reprint Edition*. Cambridge, MA: MIT Press.

Roepstorff, A., Niewöhner, J., and Beck, S. (2010). Enculturing brains through patterned practices. *Neural Networks* 23, 1051–1059. doi: 10.1016/j.neunet.2010.08.002

Sbriscia-Fioretti, B., Berchio, C., Freedberg, D., Gallese, V., and Umiltà, M. A. (2013). ERP modulation during observation of abstract paintings by Franz Kline. *PLoS One* 8:e75241. doi: 10.1371/journal.pone.0075241

Seeley, W. P. (2020). *Attentional Engines: A Perceptual Theory of the Arts*. New York, NY: Oxford University Press.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* 17, 565–573. doi: 10.1016/j.tics.2013.09.007

Seth, A. K. (2019). From unconscious inference to the beholder's share: predictive perception and human experience. *Eur. Rev.* 27, 378–410. doi: 10.1017/S1062798719000061

Shannon, C. E., and Weaver, W. (1949). *The Mathematical Theory of Communication*. Murray Hill, NJ: Lucent Technologies.

Smart, P., Heersmink, R., and Clowes, R. W. (2017). "The cognitive ecology of the internet," in *Cognition Beyond the Brain*, eds S. J. Cowley and F. Vallée-Tourangeau (Cham: Springer International Publishing), 251–282. doi: 10.1007/978-3-319-49115-8_13

Smith, M. (2012). "Empathy, expansionism, and the extended mind," in *Empathy: Philosophical and Psychological Perspectives*, eds A. Coplan and P. Goldie (Oxford: Oxford University Press), 99–117. doi: 10.1093/acprof:oso/9780199539956.003.0008

Smith, M. (2017). *Film, Art, and the Third Culture: A Naturalized Aesthetics of Film*. New York, NY: Oxford University Press.

Smith, T. J., and Henderson, J. M. (2008). Edit blindness: the relationship between attention and global change blindness in dynamic scenes. *J. Eye Mov. Res.* 2, 1–17. doi: 10.16910/jemr.2.2.6

Smith, T. J., and Martin-Portugues Santacreu, J. Y. (2017). Match-action: the role of motion and audio in creating global change blindness in film. *Media Psychol.* 20, 317–348. doi: 10.1080/15213269.2016.1160789

Smith, T. J., and Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *J. Vis.* 13, 16. doi: 10.1167/13.8.16

Stiegler, B. (2019). *The Age of Disruption. Technology and Madness (transl. by D. Ross)*. Medford, MA: Polity Press.

Stotz, K. (2010). Human nature and cognitive–developmental niche construction. *Phenom. Cog. Sci.* 9, 483–501. doi: 10.1007/s11097-010-9178-7

Sutton, J. (2010). "Exograms and interdisciplinarity: history, the extended mind, and the civilizing process," in *The Extended Mind*, ed. R. Menary (Cambridge, MA: MIT Press), 189–225. doi: 10.7551/mitpress/9780262014038.003.0009

Tajadura-Jiménez, A., Basia, M., Deroy, O., Fairhurst, M., Marquardt, N., and Bianchi-Berthouze, N. (2015). "As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ed. B. Begole (New York, NY: Association for Computing Machinery), 2943–2952.

Thompson, E. (2007). *Mind in Life*. Cambridge, MA: Harvard University Press.

Thompson, E., and Stapleton, M. (2009). Making sense of sense-making: reflections on enactive and extended mind theories. *Topoi* 28, 23–30. doi: 10.1007/s11245-008-9043-2

Thompson, E., and Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends Cogn. Sci.* 5, 418–425. doi: 10.1016/s1364-6613(00)01750-2

Tikka, P. (2010a). *Enactive Cinema: Simulatorium Eisensteinense*. Helsinki: University of Art and Design Publication Series.

Tikka, P. (2010b). Enactive media–generalising from enactive cinema. *Digit. Creativity* 21, 205–214. doi: 10.1080/14626268.2011.550028

Umiltà, M. A., Alessandra Umiltà, M., Berchio, C., Sestito, M., Freedberg, D., and Gallese, V. (2012). Abstract art and cortical motor activation: an EEG study. *Front. Hum. Neurosci.* 6:311. doi: 10.3389/fnhum.2012.00311

Van de Cruys, S., and Wagemans, J. (2011). Putting reward in art: a tentative prediction error account of visual art. *I-Perception* 2, 1035–1062. doi: 10.1068/i0466aap

Varela, F. J., Rosch, E., and Thompson, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Veissière, S., Constant, A., Ramstead, M. J. D., Friston, K. J., and Kirmayer, L. (2019). Thinking through other minds: a variational approach to cognition and culture. *Behav. Brain Sci.* 43, E90. doi: 10.1017/S0140525X19001213

Voss, C. (2011). Film experience and the formation of illusion: the spectator as 'Surrogate Body' for the cinema. (Transl.: Hediger, V., and Pollmann, I.). *Cinema J.* 50, 136–150. doi: 10.1353/cj.2011.0052

Walter, S. (2014). Situated cognition: a field guide to some open conceptual and ontological issues. *Rev. Phil. Psychol.* 5, 241–263. doi: 10.1007/s13164-013-0167-y

Wheeler, M. (2012). "In defense of extended functionalism," in *The Extended Mind*, ed. R. Menary (Cambridge, MA: MIT Press), 245–270. doi: 10.7551/mitpress/9780262014038.003.0011

Wheeler, M. (2019). "Breaking the waves: beyond parity and complementarity in the arguments for extended cognition," in *Andy Clark and His Critics*, eds M. Colombo, E. Irvine, and M. Stapleton (Oxford: Oxford University Press), 81–98.

Withagen, R., de Poel, H. J., Araújo, D., and Pepping, G. J. (2012). Affordances can invite behavior: reconsidering the relationship between affordances and agency. *New Ideas Psychol.* 30, 250–258. doi: 10.1016/j.newideapsych.2011.12.003

Wollheim, R. (1980/2015). *Art and its Objects*. Cambridge, MA: Cambridge University Press.

Zack, E., Barr, R., Gerhardstein, P., Dickerson, K., and Meltzoff, A. N. (2009). Infant imitation from television using novel touch screen technology. *Br. J. Dev. Psychol.* 27, 13–26. doi: 10.1348/026151008X334700

Zack, E., Gerhardstein, P., Meltzoff, A. N., and Barr, R. (2013). 15-month-olds' transfer of learning between touch screen and real-world displays: language cues and cognitive loads. *Scand. J. Psychol.* 54, 20–25. doi: 10.1111/sjop.12001

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 03 June 2021
doi: 10.3389/fpsyg.2021.617434

Breaking Beyond the Borders of the Brain: Self-Control as a Situated Ability

*Jumana Yahya**

Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

"I just couldn't control myself" are the infamous last words of a person that did something that they knew they should not have done. Consistent self-control is difficult to achieve, but it is also instrumental in achieving ambitious goals. Traditionally, the key to self-control has been assumed to reside in the brain. Recently, an alternative has come to light through the emergence of *situated theories of self-control*, which emphasize the causal role of specific situated factors in producing successful self-control. Some clinical interventions for motivational or impulse control disorders also incorporate certain situated factors in therapeutic practices. Despite remaining a minority, situated views and practices based on these theories have planted the seeds of a paradigm shift in the self-control literature, moving away from the idea that self-control is an ability limited to the borders of the brain. The goal of this paper is to further motivate this paradigm shift by arguing that certain situated factors show strong promise as genuine causes of successful self-control, but this potential role is too often neglected by theorists and empirical researchers. I will present empirical evidence which suggests that three specific situated factors – clenched muscles, calming or anxiety-inducing environmental cues, and social trust – exhibit a specialized effect of increasing the likelihood of successful self-control. Adopting this situated view of the ability to regulate oneself works to reinforce and emphasize the emerging trend to design therapies based on situated cognition, makes self-control more accessible and less overwhelming for laypeople and those who struggle with impulse control disorders, and opens a new avenue of empirical investigation.

Keywords: self-regulation, synchronic self-control, situated cognition, situated self-control, intracranialism, embodied self-control, extended self-control, distributed self-control

OPEN ACCESS

Edited by:
Leon De Bruin,
Radboud University Nijmegen, Netherlands

Reviewed by:
Francesco Marchi,
University of Antwerp, Belgium
Daniel Gregory,
University of Tübingen, Germany

***Correspondence:**
Jumana Yahya
jmorciglio@uos.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *14 October 2020*
Accepted: *03 May 2021*
Published: *03 June 2021*

Citation:
Yahya J (2021) Breaking Beyond the Borders of the Brain: Self-Control as a Situated Ability. Front. Psychol. 12:617434. doi: 10.3389/fpsyg.2021.617434

INTRODUCTION

"I just couldn't control myself" are the infamous last words of a person that did something that they knew they should not have done. It is exceedingly difficult to be self-controlled, especially when there are counterproductive temptations around every corner, and often, being in control of oneself is simply *too* difficult. However, for those who are capable of being consistently self-controlled, the rewards to be reaped are priceless. Self-control is instrumental for achieving ambitious goals

and those people who have mastered this ability are more successful in school (Mischel et al., 1989; Duckworth and Seligman, 2005), are better at regulating emotions (Boden and Thompson, 2015), are more likely of having a healthy body mass index (Schlam et al., 2013), are better at coping with social rejection (Ayduk et al., 2000), and are overall happier (Hofmann et al., 2014).

The impressive benefits of being self-controlled have created a demand for understanding the nature of this ability and how it can be exercised in the right sorts of ways. Traditionally, the key to understanding self-control has been assumed to reside in the brain, as evident by the persistent habit of self-control theorists constricting their scope of investigation to cognitive and neural processes. Factors that are external to the brain, such as bodily states, environmental cues, and social interactions receive a minority of attention regarding the (potential) causal roles that they play in how a self-control dilemma unfolds. However, the ever-growing number of impulse control disorders indicate that perhaps the current popular strategies for increasing self-control are not so effective and efficient.

Recently, an alternative has come to light through the emergence of several theories of self-control that go against what has become a core assumption for much of the literature: the brain is the cause of self-control. These views have their roots in situated cognition, the view that cognition depends on not only the brain, but also upon certain situated factors, including bodily states, environmental cues, and/or social interactions (Walter, 2014). Such *situated theories of self-control* emphasize the causal role of specific situated factors in producing successful self-control (e.g., Balcetis and Cole, 2009; Heath and Anderson, 2010; Vierkant, 2014). Some clinical interventions for motivational or impulse control disorders also incorporate certain situated factors in therapeutic practices, such as the focus on bodily states in mindful meditation as a therapy for addiction (Black, 2014), aggression (Singh et al., 2007), and post-traumatic stress disorder (King et al., 2013), or the focus on environmental cues in sensory rooms used to treat apathy in dementia patients (Staal et al., 2007). Despite remaining a minority, situated views, as well as the practices based on these views, have planted the seeds of a paradigm shift in the self-control literature, moving away from the idea that self-control is an ability limited to the borders of the brain.

The goal of this paper is to further motivate this paradigm shift by arguing that certain situated factors show a lot promise as genuine causes of successful self-control, but this potential role is too often neglected by theorists and empirical researchers. In order to do so, I will explain the source of contention between "traditional" and situated self-control theories in section "Setting the Stage." Then, in section "Empirical Evidence for Situated Self-Control," I will present empirical evidence which suggests that three specific situated factors – clenched muscles, calming or anxiety-inducing environmental cues, and social trust – exhibit a specialized effect of increasing the likelihood of successful self-control. Lastly, in section "Taking Stock and Moving Forward," I will take stock of the situation by briefly discussing certain implications of taking the position that self-control is situated. Adopting this view works to reinforce and emphasize the emerging trend to design therapies based on situated cognition, makes self-control more accessible and less overwhelming for laypeople and those who struggle with impulse control disorders, and opens a new avenue of empirical investigation.

SETTING THE STAGE

Theories, debates, and research pertaining to the nature of self-control comprise a large body of literature that has an extensive history and many interdisciplinary contributions. This variety contributes to the complexity and density of self-control as a concept. Similarly, situated cognition, albeit being an incredibly young concept relative to self-control, is also quite complex and dense. In order to smoothly navigate the conceptual merger between self-control and situated cognition, it is useful to review some fundamental definitions, distinctions, and terms.

In this section, I will clarify important terms and concepts that are liberally referenced throughout the remainder of the paper. First, I will provide a definition of self-control and explain certain important distinctions that are relevant to the arguments in the next section. Then, I will discuss the basic role of the brain, as well as offer some suggestions as to why the brain is assumed to bear so much of the causal burden for successful exercises of self-control. Lastly, I will present the general idea of situated cognition and how it applies to self-control.

Self-Control

It is quite an onerous task to develop a universal definition of self-control that most theorists accept without hesitation or resistance. There is a considerable amount of variation – across disciplines, as well as within – on how to define self-control[1]. For the sake of being as inclusive of the various views as possible, I will adopt the following generally broad definition:

> SELF-CONTROL refers to the ability to regulate one's own thoughts, emotions, and behaviors for the sake of achieving a particular goal(s), especially when motivational opposition is present.

This definition, albeit being non-controversial[2], is nevertheless conceptually dense and requires some further clarification before we can move on to the arguments regarding whether the brain alone is responsible for exercising this ability.

A good place to begin unpacking the proposed definition is by explaining the *function* of regulating oneself. Self-control is instrumental in achieving one's goal(s). A goal can range from being concrete and extremely specific (e.g., "I want to lose 50 pounds by Christmas") to being abstract and vague (e.g., "I want to have an attractive body"). A goal can also range from being

[1]Some of these differences are due to the fact that many self-control theorists draw a distinction between self-control and other related concepts such as willpower (Holton, 2003) and self-regulation (Fujita et al., 2018), while other theorists conceptually consolidate these into one basic ability (e.g., Sripada, 2014). I do not make these distinctions and will use these terms interchangeably.

[2]In the sense that this definition is unlikely to be accused of being too restrictive of what kinds of strategies can count as self-control (c.f. Sripada, 2014 or Fujita, 2011).

achievable in a relatively short span of time (e.g., taking an Introduction to Business course, which takes one semester) or it can require a more long-term commitment (e.g., pursuing a Master of Business degree, which takes several years). The goals which people usually care about the most are those which may be called "higher aspirations," such as improving one's social status, becoming wealthy, eliminating bad habits, being an effective leader for a large group of people (either in a professional project or in a social movement), or mastering a complex skill. Such goals tend to be more difficult to achieve than more basic desires like simply maintaining one's social status and current income, avoiding extra responsibilities, and being able to make do with the skills that one already possesses. Higher aspirations are usually formulated in an abstract or vague way (e.g., "I want to be wealthy" instead of "I want to receive a gross income of five million dollars a year by age 40"), which makes it difficult to know exactly what needs to be done in order to achieve the goal, and they are almost always long-term goals, which require extra dedication and resources to see through until the end. While self-control can certainly be useful for achieving the more concrete and short-term goals, this ability is especially beneficial for achieving the more abstract and long-term goals, as these are much more susceptible to being threatened by some form of *motivational opposition*.

Motivational opposition occurs when an agent has some reason(s) for acting in a way that is contrary to or impedes her goal(s), such as when an individual who is on a strict sugar-free diet experiences the desire to indulge in a large slice of chocolate cake. Motivational opposition also includes instances that involve some reason(s) to refrain from acting altogether, like, for example, when a very lazy individual who is passively lounging in bed has an important deadline but will not muster the energy to get up and start working[3]. When an agent experiences motivational opposition, she faces a *self-control dilemma*: she must choose between the difficult and unpleasant task of resisting the opposition, which is likely to lead to the best ultimate outcome, or she can take the easy road of succumbing to the opposition, which might feel good at the moment but will very likely lead to undesirable consequences. The agent must recognize and acknowledge that there will be negative consequences of succumbing to the opposition while, at the same time, still feeling a stronger motivational pull to succumb, as this is the crux of the dilemma[4].

To sum up thus far: the general function of self-control is to facilitate the achievement of goals, especially those goals which are more abstract and/or long-term, which also happen to be the goals which we typically care about most and hence have a strong desire to pursue. Self-control is needed in order to achieve these goals because they are vulnerable to threats by motivational opposition (i.e., the desire to do something else or the lack of desire to do the thing one is supposed to do). When opposition arises, the agent faces a self-control dilemma and must choose between resisting the opposition, which is harder but will likely result in ultimately better consequences, or succumbing to the opposition, which is easier but will likely result in ultimately worse consequences.

Succumbing to the motivational opposition with little to no resistance is essentially *weakness of will*, that is, intentionally acting contrary to one's goal(s) (McIntyre, 2006). An agent who instead chooses to resist the opposition is not necessarily self-controlled, as her efforts can either fail or succeed. An agent who tries to resist the opposition but ends up acting in a way that impedes her goal – such as the dieter who fights her craving for the chocolate cake but ends up caving into the desire by indulging in a slice – illustrates an instance of a *self-control failure*. *Successful self-control,* on the other hand, occurs when an attempt to resist some motivational opposition results in the relevant cognitive change such that the agent's corresponding behavior either promotes or, at the very least, does not impede her goal(s). Strategies or interventions which are intended to help those who are facing a self-control dilemma ideally work to increase the likelihood of successful self-control. Many philosophers who are concerned with self-control debates focus on the relationship between weakness of will and self-control by asking such questions as "how is it possible to intentionally resist some powerful temptation, which is the thing you want most right now?" (e.g., Mele, 1992; Kennett and Smith, 1996)[5].

I will take it for granted that intentional resistance against some form of motivational opposition is possible, and instead will discuss accounts of how *successful* self-control is possible. More specifically, this paper is concerned with understanding the cause(s)[6] of successfully regulating oneself and the kinds of strategies that can be implemented to ensure victory over motivational opposition. In the following part, I will explain the origin of the incredibly pervasive assumption that self-control is an ability that belongs exclusively to the brain.

[3]There are at least three different types of motivational opposition that are interchangeably discussed in the self-control literature: temptation, procrastination, and diminished motivation. Temptation refers to a competing desire, such as when an ex-smoker experiences a craving for a cigarette despite her goal to remain smoke-free. Procrastination involves a delay in pursuing one's goal, like the classic example of a college student who waits until the last minute to start her assignment even though she wants to receive a high grade in the class. Lastly, diminished motivation refers to a lack of the desire to do anything at all, including pursuing one's goals (Connor, 2013). An example of diminished motivation is clinical apathy or depression, which renders a person generally incapacitated even though a patient with this disorder can express the desire to get out of bed and live their life. In order to be inclusive, I will continue using the term "motivational opposition" to refer to all three types instead of constricting the discussions to only one type.

[4]If the agent feels a stronger motivational pull to do something that impedes her current goal(s) but does NOT recognize any negative consequences in doing so, then she is in a position to simply update her decision or revise her goal(s) regarding what is the best thing for her to do. For example, imagine a person who aims to be a vegan feels a strong desire to indulge in a juicy piece of steak, but, at the same time, she does *not* recognize or acknowledge any negative consequences of eating the meat (e.g., she does not think meat farming is unethical; or she does not believe that eating meat is bad for her health in any way). We would expect such a person to either drop being vegan as one of her goals, as she evidently does not have any reason(s) motivating her to be vegan, or to start acknowledging some negative consequence of eating animal products.

[5]This question is based on *the puzzle of synchronic self-control*. While attempting to solve this puzzle is outside of the scope of this paper, it is important to mention because this puzzle inspired many prominent theories of self-control. For an exact formulation of the puzzle and the debates that have arisen from attempted answers, see Sripada (2014) or Connor (2013).

[6]I clarify what I mean by *cause* in later sections.

The Role of the Brain and the General Neglect of Situated Factors

While there is much philosophical debate about the nature of self-control, there is significant consensus about the psychology and neurology of self-control amongst researchers and medical professionals. There are certain empirical observations regarding the importance of mindset which lead researchers to draw a connection between self-control and the brain, but emphasis of this connection likely leads to the neglect toward considering the potential role of situated factors in self-control.

There is considerable evidence that a particular cognitive state, or mindset, works to significantly increase the likelihood of successful self-control. This mindset is comprised of several related beliefs and feelings: that one is autonomous and competent (Ryan and Deci, 2000), that one's attributes are malleable rather than fixed (Burnette et al., 2013), confidence and affirmation of one's own worth (Vandellen et al., 2012), pride in one's own achievements (Tracy, 2016), and passionate determination to persevere in the face of challenges (Duckworth and Quinn, 2009); these various cognitive states contribute to one's perception of self, specifically pertaining to themes such as strength, control, and power. Taken together, these studies indicate that a specific mindset, namely, the affirmation of one's own strength, control, and power significantly increases the likelihood of successful self-control.

Based on the suggestion that a specific mindset can cause self-control, a quite common prescription for increasing self-control is to manipulate certain cognitive states; the idea is that changing the thought process changes the behavior. Consequently, the most common sorts of strategies for increasing the likelihood of self-control involve mental actions such as shifting attention (e.g., Mischel, 2014) or inhibiting recalcitrant desires (e.g., Sripada, 2014). The persistence of prescribing self-controlling strategies that consistently require some form of mental gymnastics – that is, consciously effortful mental feats – with no suggestions of how to manipulate certain bodily, environmental, or social factors is a strong indicator that the design of such strategies reflects a bias where the potential direct impact of situated factors on the success of self-control is significantly neglected.

Strategies involving shifting attention or inhibiting recalcitrant desires often recruit certain mental functions, like executive attention, inhibition, or working memory. These mental functions are correlated with certain neural areas, which happen to be located within the prefrontal cortex; the brain area that is perhaps the most associated with self-control is the ventromedial prefrontal cortex, including areas such as the orbitofrontal cortex, the lateral prefrontal cortex, and the anterior cingulate cortex (Heatherton, 2011). The relationship between the mental functions recruited for self-control and the neural correlates with which they are associated is further reinforced by the success of certain approaches that incorporate neural activity as an integral part of the therapy, such as measuring activity in the lateral prefrontal cortex to gauge cognitive training of proactive cognitive control (Berkman et al., 2014), or using amygdala activity to help implement attention bias modification to attenuate anxiety (Britton et al., 2014). So, while there certainly seems to be some sort of connection between self-control and the brain, the complexity of the brain makes it quite difficult to explain exactly what this connection amounts to (Berkman, 2018) and any tentative conclusions about this connection should be treated with caution. The emphasis that many self-control theorists place on the brain within their discussions of how it is possible to exercise this ability (e.g., Knoch and Fehr, 2007) threatens to further perpetuate negligence toward the potential role that situated factors play.

For the sake of both clarity and ease, I will call views that assume that self-control is caused only by the brain "intracranialist" positions since such views constrict this ability to the confines of the cranium. Furthermore, when I reference "the brain" or "brain-based strategies," I am referring to the cognitive processes that are consciously recruited for self-control or the strategies that rely exclusively on these processes. In the next part, I will explain the fundamental differences between a situated and an intracranialist view of self-control.

Situated Self-Control

Situated cognition is an umbrella concept which denotes any view that the mind is not constricted to the borders of the brain, but also involves some situated factors (e.g., bodily states, environmental cues, and/or social interactions) as either a *cause* or a *constituent* of cognition (Clark and Chalmers, 1998; Walter, 2014). The term situated is used very broadly and comes in many different flavors. Situated cognition includes any theories relating to the mind that can be called *embodied* (i.e., emphasis on either the causal or constituent relation between cognition and bodily states), *embedded* (i.e., emphasis on the causal relation between cognition and environmental cues), *extended* (i.e., emphasis on the constituent relation between cognition and environmental cues), *enacted* (i.e., emphasis on sense-making through interactions between bodily states, environmental cues, and social interactions), or *distributed* (i.e., emphasis on the relation between cognition and social interactions) (Walter, 2014). Situated cognition is a concept directly opposing that of intracranialism, or the view that the brain alone is responsible for cognition, which has been the dominant assumption within the cognitive sciences. Some have applied the concept of situated cognition to specific cognitive states and processes, affectivity being currently the most popular (e.g., Fuchs and Koch, 2014; Colombetti and Krueger, 2015; Colombetti, 2017). Stephan et al. (2014) nicely encompass this paradigmatic pivot with a single question: is it possible that "the brain alone can do some emoting?" One can probably pose this question for an array of different cognitive states, including cognitive abilities like self-control. When considering whether the brain alone can do some *self-controlling*, a handful of situated theories of self-control have emerged (e.g., Balcetis and Cole, 2009; Heath and Anderson, 2010; Hung and Labroo, 2011; Vierkant, 2014). Situated theories of self-control show promise for evolving our understanding and knowledge of self-control, and the practical implications alone – in terms of designing alternative

therapies for disorders of the self (Krueger and Colombetti, 2018) – should be sufficient for these views to gain significant attention. Considering that such views, unfortunately, remain the minority within the literature, it becomes important to seriously revisit this question: "can the brain alone do some self-controlling?"

The answer, as it turns out, is a bit complicated. If we take "doing some self-controlling" as ascribing causal responsibility, then one can defend a variety of claims. One can take an extreme position and argue that either the brain alone or situated factors alone can have any sort of impact on self-control. It is also possible to take a weaker position and argue that both the brain and situated factors have an impact on self-control, but the kind of impact can vary. In order to explain this distinction between different kinds of impact, I will use the word *cause* to refer to a thing that directly and consistently produces an effect, and *influence* to refer to a thing which that facilitates an effect, simply by making the surrounding conditions more favorable for the effect to take place (c.f. Sripada, 2014 for similar distinction). Considering the variety of claims that each position can defend highlights that the fundamental difference between some specific intracranialist and situated views regarding self-control can be quite nuanced. Below are five substantially different claims that can be defended by either an intracranialist or situated view of self-control:

(1) The brain causes self-control.
(2) The brain causes self-control, although situated factors can have an influence.
(3) The brain and situated factors both cause self-control.
(4) Situated factors cause self-control, although the brain could have an influence.
(5) Situated factors cause self-control.

Claims (1) and (5) represent the two most extreme positions that one can take regarding the cause of successful self-control. Claim (2) is a weaker version of an intracranialist view, whereas claim (4) is a weaker version of a situated view. It is more accurate to identify claim (3) as a situated position since ascribing causal responsibility to something outside of the cranium acts as a counterexample to intracranialism. In other words, endorsing claim (3), and thus also admitting that certain situated factors have as much causal responsibility as the brain, is incompatible with the core assumption that self-control operates only within the borders of the cranium. For these reasons, claims (1) – (5) can be assigned the following positions:

Intracranialist Views of Self-Control

(STRONG) The brain causes self-control.
(WEAK) The brain causes self-control, although situated factors can have an influence.

Situated Views of Self-Control

(WEAK) The brain and situated factors both cause self-control.
(INTERMEDIATE) Situated factors cause self-control, although the brain could have an influence.
(STRONG) Situated factors cause self-control.

In the next section, I will argue in support of the weak position of situated self-control because my goal is not to denounce the role of the brain. Rather, my aim is to emphasize the role that certain situated factors play in significantly increasing the likelihood of successful self-control, to the extent that such factors ought to be considered just as an important for self-regulation as the brain.

EMPIRICAL EVIDENCE FOR SITUATED SELF-CONTROL

The claim that certain situated factors can cause self-control has not been explicitly tested in the thorough and rigorous way that it arguably deserves. However, there is some empirical work that can shine some light on the matter. First, it is important to have a standard set of criteria for what counts as a cause, in order to be able to systematically analyze different situated factors to see which qualify as situated causes of self-control. Having an impact on self-control is by itself an insufficient criterion because too many irrelevant factors can be included. Eating ample amounts of vitamin C, for example, leads to an energetic state, but simply having energy does not guarantee success over motivational opposition, although it certainly helps. A mere influence has a general impact on self-control, whereas a bona fide cause must satisfy stricter criteria.

In this section, I will provide empirical evidence that suggests that certain situated factors have causal power in bringing about successful self-control. First, I will present a set of studies that demonstrate the causal power of a certain bodily state and briefly discuss the criteria which the investigators adopt to identify a genuine cause of self-control; namely that the factor in question must have a *specialized effect*. Then, I will present an experiment that suggests that a particular type of environmental cue can replenish self-control resources and apply these criteria to indicate that this may also be an example of a genuine situated cause of successful self-control. Finally, I will do the same thing for an example involving a particular social cue and its potential specialized effect on delay of gratification.

Bodily Cause Identifies Criteria

In recent years, with the surge in popularity of eastern philosophical ideas and practices, the concept of embodiment has gained quite a lot of attention. The harmony between mind and body is a central tenet in many current self-development practices, such as practicing yoga, breath-work, and mindful meditation, or meticulously planning one's nutrition to include as much "brain food" as the body can feasibly process. The world of clinical psychology has also joined the trend by incorporating embodiment into the design of therapies, using dance, for example, to express oneself and as a therapeutic release of energy. While the concepts of embodied cognition (e.g., Pulvermueller, 2005) and embodied affectivity (e.g., Fuchs and Koch, 2014) have received significant empirical and theoretical support (as well as their fair share of criticism), embodied self-control is a concept that has been discussed only by a small

minority (e.g., Balcetis and Cole, 2009). Can certain bodily states cause successful self-control?

The most direct evidence for the effect that certain bodily states have on successful self-control comes from a set of experiments that demonstrate that muscle tension (e.g., clenched fists or tightened calf muscles) significantly increases self-control in a variety of domains (Hung and Labroo, 2011). This set of studies aims to confirm that the physical expression of recruiting and firming willpower (e.g., clenching one's fists) also works to recruit and firm willpower. The results reveal that participants who were clenching their muscles were much more successful than their relaxed counterparts at completing an array of self-control related tasks, such as being able to withstand the discomfort of attending to unwanted stimuli, drinking large amounts of a disgusting vinegar-based "health drink," enduring physical pain for long periods of time, and making healthier food choices during snack time.

The bodily state of firming one's muscles qualifies as a legitimate cause of self-control for two reasons: (1) instead of modulating the cognitive state which then mediates the success or failure of self-control, this bodily state has a non-conscious and *direct impact* on self-control[7]; and (2) this bodily state has a *specific impact,* in that it works only to *improve* self-control, in virtue of being "inherently tied to [strengthening or] summoning willpower" (Hung and Labroo, 2011). Creating this *specialized effect* (i.e., direct and specific impact) is a crucial criterion for identifying whether some situated factor is a cause of successful self-control[8]. If the presence of some situated factor has an effect on self-control, but this effect is indirect and/or general (e.g., affirming the belief that one is autonomous and competent so that this belief improves self-control), then it is more appropriate to characterize the situated factor in question as a mere influence rather than a bona fide cause.

Unfortunately, given how young and underdeveloped the concept of situated self-control happens to be right now, there is not much additional evidence that clearly supports a causal link between various situated factors and successful self-control. While the relationships between situated factors and cognition or affectivity have received a considerable amount of empirical attention, situated self-control has yet to receive its fair share of investigation. However, being equipped with at least one criterion for identifying these situated causes makes it easier to assess other empirical studies that are not explicitly endorsing situated self-control but are nevertheless relevant. In the following part, I will apply this criterion to an example consisting of a specific type of environmental cue that replenishes self-control resources in order to propose that certain environmental cues can also be potential causes of successful self-control.

Candidates for (Environmental) Situated Causes of Self-Control

A person's immediate environment contains numerous cues that can directly affect certain cognitive states, such as the smell of lavender working to decrease stress and attenuate the perception of pain (Kim et al., 2011). Features of one's environment can also directly affect certain behaviors, like red-colored plates working to curb excessive eating (Genschow et al., 2012). If a particular environmental cue produces a specialized effect (i.e., an increase in the likelihood of successful self-control in virtue of this cue being inherently tied to strengthening willpower), then such a cue becomes an eligible candidate for being a cause of self-control. Two such eligible candidates are the calm-inducing cues found in natural environments and the anxiety-inducing cues found in urban environments.

Based on evidence that a natural environment has a restorative effect on cognitive processes (Gamble et al., 2014), one study investigates whether environment type can restore self-control resources and finds that environmental compatibility modulates this effect, in that the type of environment has to be compatible with the individual's personality (Newman and Brucks, 2016). More specifically, natural environments have a restorative effect only for personality types low in neuroticism, whereas personality types high in neuroticism experience the same restorative effect in urban environments (Newman and Brucks, 2016). The proposed reason for this effect is that processing certain environmental cues can require less attentional resources because of a sense of familiarity between the personality type and the type of cue, therefore allowing the resources to be replenished. Individuals high in neuroticism, for example, can process complex and dynamic environmental cues with less attentional effort because such individuals are more familiar or comfortable with anxiety-associated cues. This familiarity makes engaging with urban environments require less attentional resources, thus urban environments offer opportunity to recuperate and, in that sense, could be more restorative for neurotic agents. Conversely, calming environmental cues would require more attentional resources from a neurotic agent because of the lack of familiarity – such cues would have to be processed similarly to how novel cues are processed – which would prevent a state of recuperation and replenishment of resources.

So, do these specific kinds of environmental cues have a direct and specific impact on improving self-control? Well, the impact of these cues certainly seems direct in the sense that they affect self-control (resources) directly rather than modulating the cognitive state that, in turn, modulates the likelihood of self-control. In order to determine if such cues satisfy the *specific impact* condition, these cues must work to improve self-control in virtue of being inherently tied to strengthening or summoning willpower.

Currently, there is not enough empirical data to claim, with certainty, that strengthening or summoning willpower

[7]This point is corroborated especially by the second experiment in the set, where participants had to endure physical pain and those who were clenching their fists endured the pain for significantly longer than those whose fingers were kept loose and relaxed; the researchers also manipulated for belief modulation and found that clenching fists did NOT modulate any beliefs or self-perceptions, showing that firmed muscles have a direct impact on self-control.

[8]Importantly, since a cause of self-control works to increase the likelihood of successful self-control by having a direct and specific impact on successful self-control, then the brain (rightfully) qualifies as a cause of successful self-control.

inherently involves some type(s) of environmental cues. To my knowledge, such empirical investigation has not yet been conducted. It is nonetheless possible to speculate based on certain colloquial beliefs about the power of environments in providing certain advantages during competitions. In the world of competitive sports, for example, there is an idea known as the 'home team advantage', which posits that players who perform within their "home" arena, where most of their practice sessions and some of their competitions occur, are privy to an advantage over the players who are performing in this arena for the first time (Courneya and Carron, 1992; Swartz and Arce, 2014). One of the potential reasons why the home arena provides such an advantage is due to familiarity with the stable environmental cues (e.g., the layout of the arena), which makes it much easier and quicker to process the immediate environment, thus freeing up cognitive processes to focus on the matter at hand (Legaz-Arrese et al., 2013), which, in this case, is beating the competition. In this sense, certain environmental cues that comprise the "battle arena" can be construed as being inherently tied to the battle itself, such that changes in the arena directly impact the performance. While it is obvious what a home arena is within the context of sports competitions, it is much less obvious what would comprise a home arena – an environment that provides a competitive advantage based on familiarity – in a self-control dilemma. However, the concept of a home team advantage reflects the concept of environmental compatibility highlighted by Newman and Brucks (2016), in that familiarity with a specific type of cue (i.e., anxiety-inducing or calming) provides an advantage for self-control, namely, self-control resources being replenished. Following the analogy of a sports competition, it is plausible that environmental compatibility provides an advantage for an agent who is facing a self-control dilemma due to the inherent relationship between the arena (i.e., whether the agent is in an environment which contains calming or anxiety-inducing cues) and the agent herself (i.e., whether she is low or high in neuroticism, respectively).

Newman and Brucks (2016) demonstrate that calm/anxiety-inducing cues work to replenish self-control resources when the type of cue is compatible with the personality type of the agent. This observation by itself might not provide direct evidence that these specific types of cues are situated causes of self-control, since whether these cues exhibit a *specific* impact (i.e., improve self-control in virtue of being inherently tied to strengthening/summoning willpower) has not (yet) been explicitly investigated. However, as previously mentioned, the lack of explicit empirical evidence may well be due to a general lack of diligent investigation. The aim here is not to provide a convincing argument that calming/anxiety-inducing cues are undoubtedly situated causes of successful self-control, but rather to show the plausibility of identifying bona fide situated causes of successful self-control by sharing empirical evidence that strongly hints in this theoretical direction. Further empirical corroboration is needed in order to establish these cues, or other qualifying environmental factors, as situated causes of self-control. In the following part, I will discuss a certain social cue that appears to be important for an individual's willingness to delay her gratification and argue that this social cue is another plausible situated cause of successful self-control.

Candidate for (Social) Situated Cause of Self-Control

High achievers often cite the quality and depth of their social networks as one of the keys to their success. The idea that social support is a powerful tool is a key tenet of addiction recovery groups such as Alcoholics Anonymous and Narcotics Anonymous. In our modern world, people can become millionaires simply by building communities on social media platforms like Instagram, YouTube, and Facebook. It is difficult to deny the powerful effect of social factors on cognition and behavior, but can such factors qualify as genuine causes of successful self-control?

There is at least one social factor that appears to be an eligible candidate for being a situated cause of self-control: trust. Two related experiments reveal that impressions of trustworthiness affect the willingness to delay gratification (Michaelson et al., 2013). Participants were presented with vignettes and pictures of characters that vary in implicit trustworthiness (e.g., pictures of people exhibiting "untrustworthy" facial expressions) and then placed in a classic (hypothetical) delay of gratification scenario (i.e., given a choice between an immediate smaller reward or a later larger reward) with those same characters. Participants who were paired with untrustworthy characters were more likely to choose the lesser but more certain reward, whereas those paired with trustworthy characters were significantly more willing to delay their instant gratification in exchange for the larger later reward. A follow-up experiment confirmed that trust has this impact on the willingness to delay gratification irrespective of other relevant factors, such as exerting cognitive effort to regulate oneself or intentionally modulating the perception of reward (Michaelson et al., 2013).

The impression of trustworthiness has a direct impact on self-control, and this social factor works to specifically *improve* self-control by increasing the willingness to delay gratification. In considering whether trustworthiness qualifies as a situated cause of successful self-control, the question which remains to be answered asks whether trustworthiness is inherently tied to strengthening or summoning willpower. Just as with the case of environmental cues, there is yet to be conclusive evidence that trustworthiness is inherently tied to strengthening or summoning willpower, but it is possible to speculate.

While many instances of self-control dilemmas are experienced privately and thus do not contain a social dimension, it can be argued that all instances of a self-control dilemma necessarily involve trustworthiness. A key premise for such an argument is that trustworthiness applies not just to (other) social being, but also to certain non-social factors such as one's immediate environment. For example, Krueger and Colombetti (2018) argue that trustworthy access to certain affordances provided by one's immediate environment is crucial

for the regulation of affective states. Take, for example, an affordance provided by the whiteboard hanging in my office, namely, that I can use this board to write down important reminders and thus not worry about constantly keeping this information in my working memory. For this information to have an impact on my behavior (e.g., sitting down in front of my phone because it is written on the whiteboard that I have a call meeting coming up), I must trust the information written on this board. If, to push the example further, I bought this whiteboard at a joke shop and I know that any memos I write to myself are not reliable because the whiteboard changes numbers that are written on it, then seeing a call meeting reminder for a specific time written on this board will *not* motivate me to take out my phone and prepare for a meeting. Similarly, if I accidentally purchase the prank whiteboard thinking it is an average whiteboard, then I will have no reason to doubt the reliability of what is written on that board and I will sit down for my meeting at the wrong time. The point here is that throughout the different variations of this scenario, my behavior – what time I sit down to prepare for my meeting – is highly dependent on whether I perceive the whiteboard as a reliable reminder. When I trust the whiteboard, the memos correspondingly affect my behavior, but not when I perceive the whiteboard to be untrustworthy. Trustworthiness, therefore, does not necessarily apply to only people, and could very well be a factor inherently tied to strengthening or summoning willpower.

To reiterate once more, this speculation of how trust might inherently be tied with strengthening or summoning willpower is not meant to be undoubtedly convincing. Instead, the aim of this section is to suggest a viable candidate that has been shown to have a direct impact on self-control. The more viable candidates that are proposed, the more motivation there is for a paradigmatic shift of focus toward being more diligent and serious about considering situated factors as potential causes of successful self-control.

TAKING STOCK AND MOVING FORWARD

I have presented evidence that supports the claim that self-control is situated, in that certain situated factors have a direct and specific impact on improving self-control in virtue of being inherently tied to strengthening or summoning willpower. Studies that support the causal power of three situated factors (i.e., bodily state, environmental cue, and social cue) were discussed as potentially demonstrating examples of situated causes of self-control. The first set of experiments present explicit evidence for the causal power that clenched muscles exhibit over successful self-control (Hung and Labroo, 2011).

Another study provides evidence that the presence of calming or anxiety-inducing cues works to replenish self-control resources for non-neurotic or highly neurotic individuals, respectively. This study reveals a direct impact of such cues on self-control resources but does not investigate whether these cues have a specific impact of improving self-control in virtue of the cues being inherently tied to strengthening or summoning willpower. I provided some intuitive speculations of how such an inherent relationship could plausibly exist, but further empirical testing is required to establish that such a relationship indeed exists. Similarly, two related studies reveal that impressions of trustworthiness directly impact the willingness to delay gratification, but the researchers do not offer any arguments as to whether trust is inherently tied to strengthening and summoning willpower. I provided some speculations on this point as well. Admittedly, only the first example explicitly shows some situated factor exhibiting a specific impact of improving self-control in virtue of being inherently tied with strengthening or summoning willpower, but the other two examples reveal, at the very least, viable and promising candidates.

Although there is currently no demonstrative proof that these two situated factors are inherently tied with willpower, a plausible and empirically verifiable story can be told, thus contributing to the viability of their candidacy for being considered bona fide causes of self-control. It is not very surprising that older theories did not consider the potential role of situated factors in producing successful self-control, given how the popularity of situated cognition is relatively new. What is surprising, however, is how little attention the concept of situated self-control has received in contemporary research compared to much indication there is that this would be a worthwhile empirical and philosophical investigation.

One major practical benefit of unburdening the brain of sole causal responsibility for successful self-control is that exercising this ability becomes exponentially easier. Since situated causes operate non-consciously and in a reflexive-like way, the result can be achieved without conscious effort, and not having to intentionally invest conscious effort greatly reduces – if not eliminates altogether – feelings of struggle or difficulty. Delegating the work of regulating oneself to non-conscious processes thus creates an "effortless" experience. Since the anticipation of struggle or difficulty is what causes many people who face a self-control dilemma to feel too overwhelmed to attempt being self-controlled (Milyavskaya and Inzlicht, 2017), a less effortful experience can circumvent this consequence.

The goal of this paper is to make the case that empirical research concerned with self-control, as well as therapeutic interventions that are designed to treat impulse control disorders, will greatly benefit from abandoning the idea that the brain alone is causally responsible for successful self-controlling. Currently, some situated theories of self-control have already been offered and there have even been some experimental interventions that rely on situated factors to provide therapeutic benefits. However, such theories and therapies should no longer be just interesting alternatives and deserve much more theoretical and empirical attention than they have thus far received. There is so much potential for creativity, growth, and innovation for the interdisciplinary field of self-control research, and this full potential can be unleashed by simply breaking beyond the borders of the brain.

AUTHOR CONTRIBUTIONS

JY contributed equally to the writing and editing of this article.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number GRK-2185/1 (DFG Research Training Group Situated Cognition). Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer GRK-2185/1 (DFG – Graduiertenkolleg Situated Cognition). Additional support for open access fees provided by Open Access Publishing Fund of the Osnabrück University.

ACKNOWLEDGMENTS

The author would like to thank members of the Reading Club "Situated Affectivity" for incredibly helpful feedback of previous drafts. The author would also like to thank the two reviewers who have offered constructive criticisms and insightful suggestions on how to improve this manuscript.

REFERENCES

Ayduk, O., Mendoza-Denton, R., Mischel, W., Downey, G., Peake, P. K., and Rodriguez, M. (2000). Regulating the interpersonal self: strategic self-regulation for coping with rejection sensitivity. *J. Pers. Soc. Psychol.* 79, 776–792. doi: 10.1037/0022-3514.79.5.776

Balcetis, E., and Cole, S. (2009). Body in mind: the role of embodied cognition in self-regulation. *Soc. Pers. Psychol. Compass* 3, 759–774. doi: 10.1111/j.1751-9004.2009.00197.x

Berkman, E. T. (2018). "The neuroscience of self-control," in *Handbook of Self-Control in Health and Wellbeing*, eds D. de Ridder, M. Adriaanse, and K. Fujita (Abingdon-on-Thames: Routledge).

Berkman, E. T., Kahn, L. E., and Merchant, J. S. (2014). Training-induced changes in inhibitory control network activity. *J. Neurosci.* 34, 149–157. doi: 10.1523/jneurosci.3564-13.2014

Black, D. S. (2014). Mindfulness-based interventions: an antidote to suffering in the context of substance use, misuse, and addiction. *Subst. Use Misuse* 49, 487–491. doi: 10.3109/10826084.2014.860749

Boden, M. T., and Thompson, R. J. (2015). Facets of emotional awareness and associations with emotion regulation and depression. *Emotion* 15, 399–410. doi: 10.1037/emo0000057

Britton, J. C., Suway, J. G., Clementi, M. A., Fox, N. A., Pine, D. S., and Bar-Haim, Y. (2014). Neural changes with attention bias modification for anxiety: a randomized trial. *Soc. Cogn. Affect. Neurosci.* 10, 913–920. doi: 10.1093/scan/nsu141

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., and Finkel, E. J. (2013). Mind-sets matter: a meta-analytic review of implicit theories and self-regulation. *Psychol. Bull.* 139, 655–701. doi: 10.1037/a0029531

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

Colombetti, G. (2017). Enactive affectivity, extended. *Topoi* 36, 445–455. doi: 10.1007/s11245-015-9335-2

Colombetti, G., and Krueger, J. (2015). Scaffoldings of the affective mind. *Philos. Psychol.* 28, 1157–1176. doi: 10.1080/09515089.2014.976334

Connor, T. D. (2013). Self-control, willpower and the problem of diminished motivation. *Philos. Stud.* 168, 783–796. doi: 10.1007/s11098-013-0162-2

Courneya, K. S., and Carron, A. V. (1992). The home advantage in sport competitions: a literature review. *J. Sport Exerc. Psychol.* 14, 13–27. doi: 10.1123/jsep.14.1.13

Duckworth, A. L., and Quinn, P. D. (2009). Development and validation of the short grit scale (grit- s). *J. Pers. Assess.* 91, 166–174. doi: 10.1080/00223890802634290

Duckworth, A. L., and Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychol. Sci.* 16, 939–944. doi: 10.1111/j.1467-9280.2005.01641.x

Fuchs, T., and Koch, S. C. (2014). Embodied affectivity: on moving and being moved. *Front. Psychol.* 5:508. doi: 10.3389/fpsyg.2014.00508

Fujita, K. (2011). On conceptualizing self-control as more than the effortful inhibition of impulses. *Pers. Soc. Psychol. Rev.* 15, 352–366. doi: 10.1177/1088868311411165

Fujita, K., Carnevale, J. J., and Trope, Y. (2018). Understanding self-control as a whole vs. part dynamic. *Neuroethics* 11, 283–296. doi: 10.1007/s12152-016-9250-2

Gamble, K. R., Howard, J. H., and Howard, D. V. (2014). Not just scenery: viewing nature pictures improves executive attention in older adults. *Exp. Aging Res.* 40, 513–530. doi: 10.1080/0361073x.2014.956618

Genschow, O., Reutner, L., and Wänke, M. (2012). The color red reduces snack food and soft drink intake. *Appetite* 58, 699–702. doi: 10.1016/j.appet.2011.12.023

Heath, J., and Anderson, J. (2010). "Procrastination and the extended will," in *The Thief of Time: Philosophical Essays on Procrastination*, eds C. Andreou and M. White (New York, NY: Oxford University Press), 233–252. doi: 10.1093/acprof:oso/9780195376685.003.0014

Heatherton, T. F. (2011). Neuroscience of self and self-regulation. *Ann. Rev. Psychol.* 62, 363–390. doi: 10.1146/annurev.psych.121208.131616

Hofmann, W., Luhmann, M., Fisher, R. R., Vohs, K. D., and Baumeister, R. F. (2014). Yes, but are they happy? Effects of trait self-control on affective well-being and life satisfaction. *J. Pers.* 82, 265–277. doi: 10.1111/jopy.12050

Holton, R. (2003). "How is strength of will possible?," in *Weakness of Will and Practical Irrationality*, eds S. Stroud and C. Tappolet (New York, NY: Oxford University Press), 39–67. doi: 10.1093/0199257361.003.0003

Hung, I., and Labroo, A. (2011). From firm muscles to firm willpower: understanding the role of embodied cognition in self-regulation. *J. Consum. Res.* 37, 1046–1064. doi: 10.1086/657240

Kennett, J., and Smith, M. (1996). Frog and Toad lose control. *Analysis* 56, 63–73. doi: 10.1093/analys/56.2.63

Kim, S., Kim, H. J., Yeo, J. S., Hong, S. J., Lee, J. M., and Jeon, Y. (2011). The effects of lavender oil on stress, bispectral index values, and needle insertion pain in volunteers. *J. Altern. Complement. Med.* 17, 823–826. doi: 10.1089/acm.2010.0644

King, A. P., Erickson, T. M., Giardino, N. D., Favorite, T., Rauch, S. A. M., Robinson, E., et al. (2013). A Pilot study of group mindfulness-based cognitive therapy (MBCT) for combat veterans with posttraumatic stress disorder (PTSD). *Depress. Anxiety* 30, 638–645. doi: 10.1002/da.22104

Knoch, D., and Fehr, E. (2007). Resisting the power of temptations: the right prefrontal cortex and self-control. *Ann. N. Y. Acad. Sci.* 1104, 123–134. doi: 10.1196/annals.1390.004

Krueger, J., and Colombetti, G. (2018). Affective affordances and psychopathology. *Discip. Filosofiche* 18, 221–247. doi: 10.2307/j.ctv8xnhwc.14

Legaz-Arrese, A., Moliner-Urdiales, D., and Munguía-Izquierdo, D. (2013). Home advantage and sports performance: evidence, causes and psychological implications. *Univ. Psychol.* 12, 933–943.

McIntyre, A. (2006). What is wrong with weakness of will? *J. Philos.* 103, 284–311. doi: 10.5840/jphil2006103619

Mele, A. R. (1992). Akrasia, self-control, and second-order desires. *Noûs* 26, 281–302. doi: 10.2307/2215955

Michaelson, L., de la Vega, A., Chatham, C. H., and Munakata, Y. (2013). Delaying gratification depends on social trust. *Front. Psychol.* 4:355. doi: 10.3389/fpsyg.2013.00355

Milyavskaya, M., and Inzlicht, M. (2017). What's so great about self-control? Examining the importance of effortful self-control and temptation in predicting real-life depletion and goal attainment. *Soc. Psychol. Pers. Sci.* 8, 603–611. doi: 10.1177/1948550616679237

Mischel, W. (2014). *The Marshmallow Test: Understanding Self-Control and How to Master It*. London: Transworld Publishers.

Mischel, W., Shoda, Y., and Rodriguez, M. I. (1989). Delay of gratification in children. *Science* 244, 933–938.

Newman, K. P., and Brucks, M. (2016). When are natural and urban environments restorative? The impact of environmental compatibility on self-control restoration. *J. Consum. Psychol.* 26, 1–7. doi: 10.1504/ier.2020.1003 4128

Pulvermueller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582. doi: 10.1038/nrn1706

Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68–78. doi: 10.1037/0003-066x.55.1.68

Schlam, T. R., Wilson, N. L., Shoda, Y., Mischel, W., and Ayduk, O. (2013). Preschoolers' delay of gratification predicts their body mass 30 years later. *J. Pediatr.* 162, 90–93. doi: 10.1016/j.jpeds.2012.06.049

Singh, N. N., Lancioni, G. E., Winton, A. S. W., Adkins, A. D., Wahler, R. G., Sabaawi, M., et al. (2007). Individuals with mental illness can control their aggressive behavior through mindful training. *Behav. Modif.* 31, 313–328. doi: 10.1177/0145445506293585

Sripada, C. S. (2014). How is willpower possible? The puzzle of synchronic self-control and the divided mind. *Noûs* 48, 41–74. doi: 10.1111/j.1468-0068.2012. 00870.x

Staal, J. A., Sacks, A., Matheis, R., Collier, L., Calia, T., Hanif, H., et al. (2007). The effects of Snoezelen (multi-sensory behavior therapy) and psychiatric care on agitation, apathy, and activities of daily living in dementia patients on a short term geriatric psychiatric inpatient unit. *Int. J. Psychiatry Med.* 37, 357–370. doi: 10.2190/pm.37.4.a

Stephan, A., Walter, S., and Wilutzky, W. (2014). Emotions beyond brain and body. *Philos. Psychol.* 27, 65–81. doi: 10.1080/09515089.2013.828376

Swartz, T. B., and Arce, A. (2014). New insights involving the home team advantage. *Int. J. Sports Sci. Coach.* 9, 681–692. doi: 10.1260/1747-9541.9.4.681

Tracy, J. (2016). *Pride: The Secret of Success*. New York, NY: Houghton Mifflin Harcourt.

Vandellen, M., Knowles, M. L., Krusemark, E., Sabet, R. F., Campbell, W. K., McDowell, J. E., et al. (2012). Trait self-esteem moderates decreases in self-control following rejection: an information-processing account. *Eur. J. Pers.* 26, 123–132. doi: 10.1002/per.1845

Vierkant, T. (2014). Is willpower just another way of tying oneself to the mast? *Rev. Philos. Psychol.* 6, 779–790. doi: 10.1007/s13164-014-0198-z

Walter, S. (2014). Situated cognition: a field guide to some open conceptual and ontological issues. *Rev. Philos. Psychol.* 5, 241–263. doi: 10.1007/s13164-013-0167-y

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership