In: *Review of Philosophy and Psychology* (online first, March 2015).

Active Content Externalism

Holger Lyre
Philosophy Department & Center for Behavioral Brain Sciences (CBBS)
University of Magdeburg, Germany
Email: lyre@ovgu.de

January 2015

Abstract: *The aim of this paper is to scrutinize active externalism and its repercussions for externalism about mental content. I start from the claim that active externalism is a version of content externalism that follows from the extended cognition thesis as a thesis about cognitive vehicles. Various features of active content externalism are explored by comparison with the known forms of passive externalism – in particular with respect to the multiple realizability of the relevant external content-determining components and with respect to mental causation. A crucial result is that social externalism is already a version of active externalism. I conclude with a first sketch of a general account of meaning inspired by extended cognition: a use theory supplemented by a functional-role account.*

1. Preliminaries

The term "active externalism" has been coined by Andy Clark and David Chalmers in their seminal paper on "The extended mind" (1998). Unfortunately, the authors remain rather vague as to what precisely is meant by that label. In the vast majority of the literature the term "active externalism" has simply be taken as synonymous for "extended mind" or "extended cognition", a doctrine that itself has been termed by different other labels.[1] The key point is that extended cognition pertains to vehicles rather than content, while the term externalism in philosophy of mind and language is traditionally understood as a doctrine about content. What I want to show in this paper is that there is a straightforward under-

[1] Other labels are "vehicle externalism", "how-externalism" or "enabling externalism" (Hurley 1998, 2010), "wide computationalism" or "locational externalism" (Wilson 2010), "process externalism" (Keijzer & Schouten 2007), "environmentalism" (Rowlands 2003), or "transcranialism" (Adams & Aizawa 2010); cf. especially Hurley (2010).

standing of active externalism under which it might indeed be considered a new form of content externalism, albeit of a particular stripe. More precisely, I want to argue for the possibility that the doctrine of extended cognition (the term I prefer over the original "extended mind" for reasons to be given below), understood as a doctrine about possible extensions of cognitive vehicles, leads in a quite straightforward way to a substantive form of content externalism that may aptly be called "active content externalism", or "active externalism" for short. Active externalism differs from orthodox forms of content externalism in the sense that the cognitive subject can be directly influenced by the external, content-co-determining components – unlike, for instance, Putnam's natural kind externalism, where the external component, the nature of water, is passive in the sense that it has no direct impact on the subject's cognitive machinery, nor is it behaviorally relevant (as twin earth scenarios are supposed to show).

The structure of the paper is as follows. I use the first section to elaborate on the distinction between vehicles and content and active versus passive externalism (the latter being the standard form of content externalism) and on a couple of other useful terminological clarifications. In the second section some of the peculiarities of active content externalism shall be discussed, while the third section is devoted to social externalism. At the end of this section, in 3.3, we arrive at the result that, contrary to the orthodox view, social externalism must already be considered a version of active externalism. Finally, in section 4, the rarely asked question is conceived what a proper theory of meaning would look like that is in tune with both extended cognition and active content externalism. My proposal is that this should be a a use theory supplemented by a functional-role account.

1.1 Extended Cognition

A few preparatory remarks about the thesis of extended cognition (EC) are in order, before we delve into the externalism business. The EC thesis consists of the claim that our cognitive machinery comprises not only the brain (or the central neural system for that matter), but also parts of the body and the environment. It is therefore a thesis about extended cognitive vehicles, where the term "cognitive vehicle" should be understood in the sense of cognitive states, processes or activities.

As a simple example to illustrate the EC thesis and to disentangle it from embodied or embedded cognition, consider the example of a jigsaw puzzle: to assemble the puzzle, it proves to be extremely helpful to rotate the pieces physically, to test the geometric fit and

thus to solve the cognitive task of the puzzle by means of a direct physical interaction. According to the traditional view, bodily acts and manipulations of physical things are not part of the "real" cognitive goings-on which rather comprise processes of the brain such as perception, mental rotation (of the puzzle pieces) and internal evaluation (of the geometric fit). It takes two steps to get from here to extended cognition. First, the brain could use sensory and motor representations that encode, say, Gibsonian affordances and that possess a data format corresponding to the possible movements and bodily manipulations in our physical environment. This is one way to spell out the idea of embedded and embodied cognition: internal representations are formed and shaped according to our typical situatedness. Extended cognition goes a crucial step further and claims, second, that the "epistemic actions" (cf. Kirsh & Maglio, 1994) of the puzzle player not only serve to outsource internal computational processes, but that they inevitably and inextricably belong to the cognitive goings-on. There is no natural split or separation between the internal and the external; the full cognitive loop includes processes and states of the brain, the body as well as the physical embedding (here: the jigsaw puzzle).

I propose to divide the EC thesis into four sub-theses, which relate to different domains or regimes of cognitive extension:

1. Cognitive systems extend into the body.
2. Cognitive systems extend into the physical environment.
3. Cognitive systems extend into the `informational' environment.
4. Cognitive systems extend into the social environment.

The four domains of extended cognition are roughly motivated by and related to the ideas of embodied cognition (domain 1), embeddedness and dynamicism (domain 2), social cognition (domain 4) and situated cognition in general (cf. Robbins & Aydede 2008 for an overview). The different domains already indicate that a full list of possible extensions of cognition will be quite heterogeneous. Cognitive states and processes may include states and processes of the body, as for instance certain mechanisms of embodiment or proprioperception (domain 1), as well as states and processes of (parts of) the physical environment into which a cognitive system is embedded and to which it is dynamically coupled, such as sensorimotor loops (domain 2). They may also include states and processes of what might be called the 'informational environment' – roughly that part of the physical environment that is `informationally or computationally laden'. It includes notebooks, computers, Internet, smartphones and the like (domain 3). Finally, cognitive extension may include couplings to other minds or to social and language communities (domain 4).

In order to avoid rampant extension or cognitive bloat, it is important to specify restricting and limiting coupling criteria that must be met by the external components of a cognitive system. As already indicated by Clark and Chalmers, such criteria will roughly comprise accessibility, robustness, and reliability of the external components (and they must be spelled out in detail for all domains – a task still open for EC proponents). Certainly, such criteria are not absolute and categorical. Whether states or processes outside the boundary of skin and skull can be counted as cognitive extensions or extended vehicles is a matter of degree. Therefore, it seems to be an inevitable consequence of the very conception of EC that extended cognitive systems don't possess sharp boundaries.

1.2 Externalism

Externalism about mental content consists in the claim that the content of our mental states is at least in part determined or individuated by factors outside the brain. In other words, content supervenes on more than just internal cognitive vehicles, whereby "internal" usually refers to neural states and processes. Content externalism – in its three predominant flavors of natural kind, historical and social externalism – is generally understood in terms of passive externalism. Passive externalism is the view that the cognitive system is passive concerning the external factors that co-determine content. One can construe this either as *genitivus subjectivus* or *genitivus objectivus*: the external factors are passive in that they don't influence the cognitive system's machinery, and the system is passive concerning the external factors in that the system cannot influence the latter. Another way to spell this out is by saying that the factors in question are distal rather than proximal.

Clark and Chalmers introduce the notion of *active externalism* as an alternative to passive externalism. It remains, however, largely unclear in their paper whether active externalism pertains to vehicles or to contents. Susan Hurley (1998, 2010) argues that we should distinguish vehicle externalism from content externalism, and many have accordingly understood active externalism to be a form of vehicle externalism. But the upshot of such an understanding is that the "extended mind" and "active externalism" are just two different terms for the same doctrine. Indeed, Clark and Chalmers do not agree on this point. While Clark, in subsequent papers, restricts the extended mind thesis to a thesis about vehicles[2] and tends to avoid the term "active externalism" altogether, Chalmers acknowledges ex-

[2] „The Extended Mind hypothesis is really a hypothesis about extended vehicles, vehicles that may be distributed across brain, body and world. We conflate vehicles and contents, as Dennett (1991) and Hurley (1998) stress, at our philosophical and scientific peril" (Clark, 2005, fn. 1).

plicitly that the discussion of active externalism in their extended mind paper refers to beliefs and, hence, to *bona fide* mental contents (cf. Chalmers, 2008, and private communication).

This gives us a first hint as to how to distinguish things conceptually. The extended cognition thesis as such is not a thesis about about contents but about vehicles. It is, in Hurley's words, a version of vehicle externalism. This is probably also a tacit reason why many of the more recent authors in the literature, including Andy Clark (cf. his 2008), prefer to talk about extended cognition rather than extended mind: to avoid connotations of mind and content. I will follow this convention in this paper. However, since "externalism" is generally used as a term that refers to content, I also suggest to avoid the use of the term "vehicle externalism", and to understand the term "externalism" exclusively in the sense of "content externalism". So far, this is merely a terminological point. But we encounter a substantial philosophical thesis in the claim that active externalism is really a version of content externalism. This is the initial thesis of the present paper. A straightforward way to arrive at such a thesis is by the widespread assumption that content supervenes on vehicles. From this perspective, active externalism is a version of content externalism that is virtually implied by the extended cognition thesis as a thesis about extended vehicles. More specifically, on this view, the thesis of active externalism pertains to content that supervenes on extended vehicles.

There are three strands of theories of meaning that are directly tied to versions of content externalism. Causal theories of reference go together with Putnam's classic form of natural kind externalism; teleosemantics comes with historical externalism. While in the former case the meaning of a natural kind term like "water" is fixed by a causal reference chain, in the case of teleosemantics it is the entire adaptive history of the indicator function of a term that determines its meaning. Both causal theories and teleosemantics fall under the broad scope of realist theories of meaning. A third strand, to be discussed more further in section 3, consists of use theories of meaning (in a roughly Wittgensteinian spirit) and is tied to social externalism. What is important in the context of our considerations is that all three versions of externalism seem to share a common feature: the external components that determine content are distal and therefore not behaviorally relevant. Conversely, the cognitive subjects have no direct influence on these external components. We may conclude that all three versions of externalism are versions of *passive externalism* (we shall see below, however, that this conclusion is too quick). By way of contrast, extended cognition leads to a new form of *active externalism*. This shall be elaborated in what follows.

2 Active Externalism

2.1 From Extended Cognition to Active Content Externalism

Consider the following example: when asked whether they know what the time is many people react with an explicit "yes", look at their watch only afterwards, and finally give a response. One might dismiss this as mere elliptical speech, but perhaps the example indicates how self-evident and natural it has become for us to wear a watch as a constant personal companion. Taking the statement literally, many people attribute themselves a knowledge state which they couldn't attribute themselves without the watch. From an EC perspective this means two things: first, the legitimate cognitive subject is the enlarged system "S + clock". In greater generality and a bit more formal: S enlarges to S* = S + E, where "S" is the original cognitive system, "E" is an external component, and "+" indicates the composition of S and E into a bigger, integrated whole. Secondly, it is the extended system S* rather than S to which a knowledge state should be attributed. Only S* 'knows the time'.

The example is intended to show how EC together with the content-vehicle supervenience assumption directly implies a version of content externalism. The content of the knowledge state depends in part on factors outside the brain (here: the watch as an external cognitive vehicle). EC leads to an extended vehicle supervenience base for content. Naturally, the vehicles of such extensions can be quite heterogeneous. But even if extended cognitive systems are larger than traditionally conceived, they are nevertheless integrated wholes. This is tantamount to saying that the extended content-determining components are of proximal and not of distal nature. And this gives us an immediate reason why a form of externalism that is implied by the thesis of EC is active rather than passive. In the former case, in contrast to the latter, the relevant external components and factors can influence the cognitive system, and the system is likewise able to influence the external factors. Thus, I shall understand the *thesis of active externalism as the claim that content not only depends on internal states and processes of the cognitive system but also on external components (extended vehicles consisting of bodily or environmental states or processes), such that a variation of these external components is of functional or behavioral relevance*.

The active nature of this new form of content externalism can easily be seen from our example. It depends (in part) on the states of the watch whether we may attribute S a true or a false knowledge state, or even a knowledge state at all. S's watch might be wrong, which means that S has a false belief about the present time. Or, S's watch might be broken,

which means that S has a no knowledge at all about the time. Obviously, in all cases the particular knowledge content is, in principle, behaviorally relevant (S might miss the bus, for instance).

This active character was also vividly illustrated by Clark und Chalmers in their well-known twin earth scenario about Otto, the Alzheimer patient. Twin earth scenarios are frequently used to analyse the implications of the particular versions of externalism. Patient Otto manages his daily life by storing all important data and information in a notebook which he uses as his permanent companion. He and Inga want to meet at the Museum of Modern Art in 53rd street. Since Otto is unable to memorize the MOMA's address, he consults his notebook and finally meets Inga. Similarly, on twin earth Twotto wants to meet Twinga. But Twotto's notebook mistakenly diplays 51st street as MOMA's address. So Twotto won't meet Twinga. This shows the active nature of active content externalism: a variation of the external component of the cognitive system, in this case the notebook entry, may change mental content in a behaviorally relevant manner.

Two further points about active externalism are worth mentioning. First, its active nature seems to be very promising from a naturalist point of view. It has always been a point of controversy whether and how a passive externalist construal of content, that is a construal of content as behaviorally irrelevant, is at all convincing. On the active externalist construal of content, no such qualms occur, as content has behavioral relevance. Second, active externalism is not only different from the passive versions of externalism, it also seems to be logically independent. One can be a staunch active externalist and at the same time endorse, say, natural kind externalism. Active and passive externalism are independent, orthogonal forms of externalism. One might wonder whether a reason for this is that, unlike versions of passive externalism, active externalism is not (yet) linked to any particular theory of meaning. I will follow up on this point in the last section, but shall examine some interesting features of active externalism first.

2.2 Active Externalism, Multirealization, Mental Causation and Extended Internalism

Mental content is usually considered to be multiply realizable. The general idea is that higher-level properties possess multiple lower-level realizers or instantiations. Under the assumption that the realizers are 'drastically heterogeneous', no identities between higher-level and lower-level types can be established. This means that higher-order generalizations or laws cannot be type-reduced to lower-level laws (cf. Lyre 2009). Let us consider

the case of passive externalism: under which conditions are the external content-fixing components themselves multiply realizable?[3] Start with natural kind externalism. According to this thesis the content of my water-thoughts is fixed by the essence of water – that is, by its assumed nature of water as H_2O. But of course, the essence itself is not multiply realizable. Under the presupposition of a causal theory of reference, only the system-internal, narrow content is multiply realizable. Put this way, multiple realizability becomes a defining criterion of narrow content.

Now consider historical externalism. Teleosemantics doesn't allow for the multiple realizability of external components either. This might at first sound surprising. Isn't it possible that one and the same content is produced by different adaptive histories? This must even be assumed: my personal learning history of a specific word or phrase is certainly different from my neighbour's history, but nevertheless many if not most of the ordinary words, phrases, and sentences in a language will be understood by all competent speakers of the language in basically the same way. So it seems that the external historical component is multiply realizable. But this is a misunderstanding. From a rigorous teleosemantic point of view contents that are individuated by different histories can never be identical. For what individuates content is the ultimate evolutionary teleofunction of a certain mechanism, not its proximate causal function. The production of functionally equivalent mechanisms by evolution (equivalent with respect to proximate functions) doesn't guarantee the equivalence of the ultimate, distal teleohistories. On the contrary, a difference in external historical teleocomponents implies a difference in content, since content is individuated by such histories. To put it the other way round: if teleohistories were multiply realizable they couldn't serve to individuate contents uniquely (undermining the whole project of teleosemantics). The functional equivalence of speakers in terms of their behavior does not imply that the contents of the speakers' thoughts driving the behavior are the same. Teleocontent isn't multiply realizable; multiple realizability applies to proximate functions, not to ultimate teleofunctions.

Things are different when it comes to active content externalism. The general motivation of extended cognition is strongly supported by functionalism, it is therefore straightforward to assume that the external components of active externalism are multiply realizable. The

[3] This is a rarely posed question in the literature (if it has been considered elsewhere at all). Note that whether or not the external components of passive externalism are multiply realizable should not be confused with the possibility that content is multiply realizable in the traditional sense that the internal vehicles of a cognitive system may vary (by exchanging vehicle tokens into functionally equivalent tokens) while the content remains unchanged. (I like to thank an anonymous reviewer for his request for clarification.)

extension of my mathematical capacities can be reached by several, functionally equivalent means, for instance by either using an abacus or a pocket calculator.[4] And Otto's notebook may very well be replaced by a copy. Vehicles of cognitive extension are multiply realizable.

Another interesting feature of active externalism concerns mental causation. Cognitive agency is what constitutes humans. It is the *content* of our mental states that provides the basis for our ability to act. This is the basic idea of mental causation. However, there seems to be a strange tension between mental causation and externalism. Indeed, how can mental contents be causally efficacious if the contents themselves depend on distal external factors? The problem is that causation is generally regarded as local, so externalism seems to be in conflict with local psychophysical supervenience (cf. Jacob 1993 for an overview as well as an interesting interpretivist position). But this applies to passive externalism only. Active externalism refers to proximal, directly and locally coupled external components. Variations in such components – a different notebook entry for instance – are behaviorally relevant. As Clark and Chalmers put it:

> *"the human organism is linked with an external entity in a two-way interaction, creating a coupled system that can be seen as a cognitive system in its own right. All the components in the system play an active causal role, and they jointly govern behaviour in the same sort of way that cognition usually does. If we remove the external component the system's behavioural competence will drop, just as it would if we removed part[s] of its brain."* (Clark and Chalmers 1998, 8-9)

Because of the active nature of the extended vehicles, active externalism is in no conflict with mental causation, it just enlarges the behaviorally relevant system. It is, again, the extended system `Otto plus notebook´ which has a belief about the MOMA to be located in the 53rd street and which acts accordingly. The belief has behavioral consequences, but the system acts as a `locally extended´ system, and not as a non-local (natural kind externalism) or transtemporal system (historical externalism).[5] Of course, the active nature of active

[4] Of course, this statement holds only insofar as abacus and pocket calculator *really are* functionally equivalent. But perhaps they aren't. Modern pocket calculators do all sorts of advanced things (they allow, for instance, to display functional graphs etc.). So, on the face of it, the expressive powers of both tools are rather different. For the sake of the argument and to secure functional equivalence simply by fiat, let us consider them as nothing but instantiations of one and the same Turing machine.

[5] A word of clarification about the pairs proximal/distal and active/passive (as again, pointed out by an anonymous reviewer): The causal theory of reference aims to understand the reference relation in terms of a causal chain between the cognitive subject and the object of reference. This causal chain, however, can be very complex and long-range. I might meaningfully talk and think about Komodo dragons although I've never been in direct physical contact with any one of them. I could have seen them in TV only. The chain might even go back into my causal past, which is the reason why we can meaningfully

externalism has to do with the "active causal role" of the extended vehicles. Most authors stop at this point and characterize extended cognition as a form of vehicle externalism. However, by assuming content-vehicle supervenience we may likewise insinuate a new form of content externalism (again: Clark and Chalmers, in the passages just quoted as well as in their whole paper, are vague about the vehicle-content distinction and therefore leave this open). Content that supervenes on extended vehicles can generally be considered as behaviorally relevant. The causally relevant system is a locally extended system, and there is therefore no tension between active externalism and mental causation.

This characterization may raise suspicion in some readers. On this view, isn't active externalism simply internalism in disguise? Mental contents simply supervene on the internal states of the extended cognitive vehicle system. Accordingly, it would seem that active externalism is no externalism at all, it is good old fashioned internalism – *extended internalism*, as it were. This is, in a sense, a legitimate description. We may indeed consider active externalism to be a form of extended internalism, but this is merely a terminological point. The crucial point is that, under the supposition of extended cognition, the extended system becomes an ever varying, context-, situation- and task-dependend system. Cognitive systems aren't the neat, well-defined and well-confined systems the internalist tradition mistook them for. They are highly situated and, hence, variable systems. This does not mean that there is no boundary between a cognitive system and its environment, but it does mean that the boundary is never fixed. Parts of the world become parts of the system and lose this status again at a later point. Contents are individuated in relation to these external parts. Content is therefore a relational feature of some assumed core system and the various situation-dependent external components. And even the idea of such a core system has to be questioned (though I won't delve into this here). It is this radical revision of the tradi-

---

talk and think about Julius Caesar or other historical figures and events. It is this long-range character that is captured by the term 'distal'. How does this relate to the characterization of physical or natural kind externalism as passive or behaviorally irrelevant? Of course, direct causal contact with a Komodo dragon will most certainly be behaviorally relevant. But that is of course not the point. According to the causal theory of reference the causal chain serves to individuate *content*. So whether I'm in direct or rather indirect (long-range) causal contact with a Komodo dragon – that itself doesn't change the reference of my Komodo dragon thoughts. Nor am I able to change the Komodo dragon kind. This classifies natural kind externalism as passive. See again Clark and Chalmers:

"In these cases, the relevant external features are passive. Because of their distal nature, they play no role in driving the cognitive process in the here-and-now. This is reflected by the fact that the actions performed by me and my twin are physically indistinguishable, despite our external differences. In the cases we describe, by contrast, the relevant external features are active, playing a crucial role in the here-and-now. Because they are coupled with the human organism, they have a direct impact on the organism and on its behaviour. In these cases, the relevant parts of the world are in the loop, not dangling at the other end of a long causal chain. Concentrating on this sort of coupling leads us to an active externalism, as opposed to the passive externalism of Putnam and Burge." (Clark and Chalmers 1998, 9)

tional internalist view of fixed cognitive systems, which justifies the term "active externalism". Others might as well speak of "extended internalism" – nothing hinges on that.

3. Social Cognitive Extension and Active Social Externalism

3.1 Social Cognitive Extension and Language

Cognitive subjects can extend into the social environment, that was already suggested in the beginning. Kosslyn (2006) speaks about „social prosthetic systems: [...] in such systems, other people serve as prosthetic devices, filling in for lacks in an individual's cognitive or emotional abilities." Consider the familiar example of a longterm couple where both partners help each other to memorize things. In such cases (the requirements of accessibility, robustness and reliability assumed to be fulfilled), the social partner serves as an external aid – by finishing each other's sentences for instance – or even as an external memory. The role of the partner's memory as a social external component is fully analogous to the role of Otto's notebook. And here again, content supervenes on internal plus external vehicles. Suppose Otto uses his wife Mary as a constant source to get reminded about his heart pills in the morning. Then Otto-plus-relevant-parts-of-Mary's-memory has a belief that the pills should be taken. And obviously, this belief is behaviorally relevant. It plays an active role in Otto's life. If the external Mary-memory-component fails (for whatever reasons), this has immediate behavioral consequences.

In the Otto-Mary case as well as in many of the more advanced cases of social cognitive extension, language plays a crucial role as the decisive link between the internal system and the external social component(s). Language itself is a social practice. From the point of view of extended cogniton, a broad new picture emerges: We, as language speaking subjects, are spread into the language community that makes up our social world. The functional roles that our linguistic expressions play in the social world do in part already belong to our cognitive make-up as socially extended cognitive subjects.[6]

[6] To be sure, language is but one of many social practices. In the present context I will restrict myself to talk about linguistic social practices, but extended social cognition comprises various forms of social interactions with language being only one of them. The space of social interactions or "we-space", as Krueger (2010) has rightly pointed out, comprises of various interactive forms of "we-space management", such as gesture, touch, facial and whole-body expressions. Since such processes are driven and (partially) constituted by environmental scaffolding, we find them among the vehicles of cognitive extension into the social environment. And this is not only true of these basic mechanisms, but also of the more advanced processes and mechanisms of social cognition: from joint attention and behavior read-

If we supplement this picture with the content-vehicle supervenience assumption, we approach in fact an externalist view of language that's already familiar: social externalism, another much-noticed form of content externalism. Linguistic meanings and content supervene on the use behavior and functional roles in the language community. What extended cognition (in the social domain) adds to the picture of social externalism, a picture on the level of content, is a picture of an underlying vehicle supervenience base of socially extended cognitive systems, where some of the functional roles of language use in the social environment should be viewed as being already part of the socially extended system. This differs from the traditional picture of social externalism where speakers are considered as non-extended cognitive subjects – with language use behavior as separate and external. Since, moreover, the link between social content externalism and speaking and thinking cognitive subjects is generally provided by use theories of meaning, we are now facing a theoretical amalgam of three ingredients: (1) a theory of socially extended cognitive systems, (2) social externalism, and (3) use theories of meaning. The first is a pure vehicle, the second a pure content theory, while the third provides us with a link between content and vehicles.

The connection between (1) and (2) consists basically in the claim that linguistic meanings and content supervene on socially extended rather than non-extended cognitive systems together with the functional roles of language use. This claim, so far only hinted at here, translates the extended cognition project into a research program that scrutinizes questions of the following sort: What "parts of the functional roles of our language use" (my phrasing above) should be considered as "being already part of the socially extended system", and what are the circumstances and qualifying criteria for this? A full evaluation of this point also leads to the question, what the appropriate theory of meaning is that is in line with extended cognition. Clearly, this task goes beyond the scope of the present paper. Nevertheless, in the final section 4 some first steps should be taken into this direction. This will in particular include a proposal of a more refined distinction between different sorts of "functional roles".

First, however, we should follow the general line of our paper that aims at a characteriza-

ing over shared intentionality up to full-blown mindreading. I've addressed some of these points in a forthcoming paper, where I propose a new understanding of the mechanisms of shared intentionality as constituting some of the most decisive coupling mechanisms of extended social cognition. The key idea is that Bratman's well-known conditions of shared intentionality and cooperation can be understood as coupling conditions in the sense of section 1.1 (it is, I believe, not necessary to buy into Searle-Tollefson-like intentions-in-action, as critically discussed by Blomberg 2011; see Lyre, forthcoming).

tion of active content externalism. We have already compared active externalism with natural kind and historical externalism as two widely regarded forms of passive externalism. In the same spirit we shall now have a closer look at social externalism (2) and its connection to (3), use theories of meaning. This will be done in the following sub-section. We shall eventually see, in sub-section 3.3, that contrary to the orthodox view, social externalism turns out as a form of active rather than passive externalism.

3.2 Social Externalism

Social externalism is associated with use theories of meaning. The key idea is famously captured by Wittgenstein in his 'Philosophical Investigations': *"For a large class of cases — though not for all — in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language" (Wittgenstein, 1953, § 43)*. On this view, content supervenes on the (relevant parts of the) practices of language use in a speaker community. Corresponding twin earth scenarios were effectively already construed by Wittgenstein himself (cf. Child 2006 for an excellent discussion). Here is one of them:

> *"It is, of course, imaginable that two people belonging to a tribe unacquainted with games should sit at a chess-board and go through the moves of a game of chess; and even with all the appropriate mental accompaniments. And if we were to see it we should say they were playing chess. But now imagine a game of chess translated according to certain rules into a series of actions which we do not ordinarily associate with a game – say into yells and stamping of feet. And now suppose those two people to yell and stamp instead of playing the form of chess that we are used to; and this in such a way that their procedure is translatable by suitable rules into a game of chess. Should we still be inclined to say they were playing a game? What right would one have to say so?"* (Wittgenstein, 1953, § 200).

In Wittgenstein's view, we are neither entitled to regard the behavior of the two people as chess playing in the more obvious case of their yelling and stamping of feet, nor in the less obvious case of their externally correct behavior. In both cases, the two people are wholly unacquainted with games – the practice of playing games is completely unknown to their society. But without such a practice, without the rules of chess playing being instantiated by a practice, there can be no question of playing chess even if the behavior of the supposed players perfectly resembles the behavior of real chess players in our society. Rule following supervenes on social context. And this is true, according to Wittgenstein, even if the people "*go through the moves of a game of chess ... with all the appropriate mental accompani-*

ments". Mental content is thus also dependent on the right social context, that is on language use and practice. Content supervenes on social context and linguistic practice. Wittgenstein adopts a form of social externalism.

A more explicit and widely regarded twin earth scenario for social externalism is discussed by Tyler Burge (1979). The scenario concerns Oscar's pains in his thigh, which make him believe that he has arthritis. Arthritis, however, is a condition of the joints only, so Oscar has a false belief. On twin earth, on the other hand, "arthritis" is used such that it includes rheumatoid ailments not just of the joints but of the thighs, too. So Twoscar's belief that he has arthritis is true; meanings are co-determined by the speaker's community. Again, this is an instantiation of social externalism.

Let us, ever so briefly, examine particular aspects of social externalism under the keywords misrepresentation, normativity and multiple realizability. One of the well-known problems of causal theories of meaning is the problem of misrepresentation. How do use theories fare in this respect? As the now already classical example goes, according to causal theories of meaning, cow-representations are caused by cow-tokens only. Sometimes, however, we mistake cows with horses. A crude causal theory cannot distinguish such a case of misrepresentation from the case of representing the disjunction of something being either a cow or a horse. The crude symmetric causal condition may therefore be replaced by an antisymmetric, counter-factual condition: if cows don't cause cow-representations, then horses don't cause them either. It is, however, possible that horses don't cause cow-representations, but cows do. Disjunctive representations of cows-or-horses are asymmetrically dependent on representations of cows. Hence, misrepresentations are parasitic on successful representations. But what is a successful representation? To tell a successful representation from a misrepresentation, or successful functioning from malfunction, normative criteria are needed. It seems that naturalistic accounts do not have the required resources to capture the normative dimension of notions like representation and function.

Wittgenstein anticipates these questions in his famous rule-following considerations. Rules and norms are socially and publicly taught and sanctioned; their application does not hinge on pre-existing mental items, algorithms or built-in functions (of mechanical devices or computers). The crucial question is whether there is any fact of the matter that determines the application of a rule. One way to deal with this is to adopt a sceptical stance. Presumably Wittgenstein himself did not endorse a sceptical view (admittedly a controversial matter of exegesis), but such a view is in fact more in tune with an elaborate use theoretic account of meaning. It would still allow for a, as it were, 'thin' notion of normativity

according to which rules and, by extension, meanings are determined to be as fine-grained as the practices in the speaker community allow them to be. However, it would not allow for a 'thick' notion of genuine normativity, according to which rules and meanings are uniquely determined. Scrutinizing these questions at length is unfortunately beyond the scope of this paper.

In section 2.2, the multiple realizability of external components was discussed. We found that external components posited by active externalism but not those posited by natural kind and historical externalism are multiply realizable. But we had not yet considered social externalism. Here, contents supervene on language usage in speaker communities. But language use, rule-following practices and patterns of behavior are very well multiply realizable. It is, in principle, possible that different language communities develop the same practices and patterns of use (although this is of course highly unlikely). A particular game, say soccer, could have been invented independently in different communities. Similarly, a society X could have developed the same patterns of behavior and practices with respect to using a specific word in the X-language than society Y with respect to using a specific Y-word (although, again, this is not very likely). The observation that the external components posited by use theories, but not those posited by causal theories and by teleosemantics, are multiply realizable, raises the suspicion that there might be a deeper connection between social externalism and active externalism.

3.3 Active Social Externalism

In Burge's twin earth scenario Otto has a false belief, since arthritis really means an ailment of the joints but not of the thighs. Arthritis is *used* such that it refers to ailments of the joints only. According to use theories of meaning, meaning is determined by the language community. Meaning supervenes on the uses and practices in a language community. This is the *externalist* component of use theories of meaning. And clearly, since natural language communities generally comprise millions of speakers, each single speaker is unable to change the overall usage of words. This is what makes us think that social externalism is a form of *passive* externalism.

But now consider a simple thought experiment (cf. Lyre 2010). Suppose that the language community is far smaller. Perhaps, it consists of just a few speakers. Think of a gang of youths with their own slang. In that case, of course, each single speaker, the gang leaders in particular, will actively contribute to the overall usage of the group's idiom. We may

conclude that a use theoretic externalism is an *effectively* passive externalism only; that is an externalism which is in effect passive for the practically relevant cases of natural language communities. But strictly speaking it is an active externalism. The degree of passiveness depends basically on the size of the language community (though not solely). Social externalism is an active externalism, insofar and to the degree as to which the external component, the language community, is proximal. But it is active externalism in disguise, as it were, since the external component gradually becomes distal as the community increases and, inversely, the single speaker's influence on the community decreases. In the case of natural languages the community as such is of course no longer part of one's direct social interactions. Nevertheless, social externalism is strictly speaking a sub-variant of active externalism.[7]

4. Outlining a Functional-Role Semantics as a Semantics for Active Externalism

Use theories are associated with social externalism, and insofar as social externalism appears as a sub-variant of active externalism, it is to be expected that active externalism must generally be connected with some broader, still to be spelled out theory of meaning, which includes the basic lessons of a use theory, but which is supplemented with another semantics account. A natural guess, especially in the light of our preliminary considerations in 3.1, is that this will be a functional-role semantics (FRS). In the following I will give a first sketch of what I consider to be the natural framework for an EC inspired semantics. That sketch consists basically in a clarification of what types of (extended) functional roles are involved. It goes without saying that a deeper scrutiny leads to a major project that goes far beyond the scope of a single paper.

For our purposes, a broad understanding of FRS will suffice. I will not distinguish between the many different versions of FRS on the market, such as conceptual, inferential, computational, or causal role semantics (though such distinctions are of course important from a

[7] As Dave Chalmers has pointed out in correspondence, later reprints of the 1998 Analysis paper as well as the web version include further footnotes, some of which touch upon our point. In particular: '*Might this sort of reasoning also allow something like Burge's extended 'arthritis' beliefs? After all, I might always defer to my doctor in taking relevant actions concerning my disease. Perhaps so, but there are some clear differences. For example, any extended beliefs would be grounded in an existing active relationship with the doctor, rather than in a historical relationship to a language community. And on the current analysis, my deference to the doctor would tend to yield something like a true belief that I have some other disease in my thigh, rather than the false belief that I have arthritis there. On the other hand, if I used medical experts solely as terminological consultants, the results of Burge's analysis might be mirrored.'* (cf. footnote 9 of 'extended' reprints of Clark & Chalmers (1998) in Chalmers (2002, 643-651), Clark (2008, 220-232), or online: *philpapers.org/rec/CLATEM*).

certain point onwards). According to FRS, the meaning of a linguistic expression as well as the content of a mental representation is constituted by the role the expression or representation plays. In this respect, FRS is sometimes regarded as an extension of use theories of meaning. Use theories consider linguistic meaning to be constituted by language use and speakers' behavior or, in other words, by the functional roles a speaker's utterance or expression plays in a language game. From this point of view, a use theory of meaning is a functional role account concerning the functional roles of speakers' language-use behavior (on this point see especially Block 1998).

On the other hand, FRS can be interpreted as concerning the functional roles of the internal causal goings-on of the constituents, that is the vehicles, of a cognitive system. Such an FRS account is compatible with a purely internalist account of representational content. The crucial point is, of course, to bring the two accounts together. This can be achieved by adopting an extended view of cognition. On this view, internal neural goings-on affect via efferent neural activity bodily behavior, which in turn affects the environment via the interaction between body and environment. Hence, functional roles extend beyond the internal neural system into body and environment.

FRS can accommodate this extension, but a few amendments must be made compared to traditional FRS conceptions. Block (1998) introduces the useful distinction between short-arm and long-arm versions of functional roles. While short-arm roles stop at the traditional boundary of skin and skull, long-arm functional roles extend beyond that boundary. Short-arm functional roles comprise what is usually known as narrow content. As we will see in a moment, caution should be exercised as far as long-arm roles are concerned. They are associated with traditional externalist conceptions and comprise what is usually known as wide content. Long-arm functional roles may, for instance, include the reference fixing causal chains that are central to causal theories of meaning.

From an EC perspective, the short-arm/long-arm distinction must be supplemented by 'extended short-arm' and 'extended long-arm' functional roles. Extended short-arm functional roles comprise all functionally relevant causal interactions and processes of the extended cognitive system (which of course, from the EC point of view, just *is* the proper cognitive system). The extended cognitive system, again, includes the traditionally conceived cognitive system, the neural system, plus all suitably coupled external components such as bodily parts and parts of the physical, informational and social environmental embedding (as far as they are relevant for a given cognitive task). Extended short-arm functional roles comprise the functional roles in a network that is constituted by this extended set of cogni-

tive vehicles. We may, for reasons of brevity and simplicity, just call them 'extended functional roles' – and suppress the 'short-arm', since extended long-arm functional roles won't be part of a straightforward EC inspired semantics. Why? Extended long-arm functional roles transcend the boundary not just of skin and skull, but also of the extended cognitive system. They are of interest for those who want to combine EC with a version of a purely passive externalism (i.e. natural kind or historical externalism). But passive externalism is unsatisfactory, if not mysterious, insofar as it must assume behaviorally irrelevant content-determining external factors (as outlined in section 2). More importantly, from a strict EC point of view, it is straightforward not to combine active with passive externalism, but to stick with a purely active version instead. Under this supposition, all functional roles are extended short-arm or just 'extended functional roles'.

Moreover and as already mentioned in the beginning, cognitive extension occurs in different domains. Cognitive states and processes may extend into the body, the physical, informational, and social environment. We may accordingly distinguish between bodily, physically, informationally, or socially extended functional roles which are constituted by the corresponding networks comprised of vehicles in the various domains. In the present paper particular attention has been paid to the social environment. Language use and speakers' behavior can be reconstructed as *socially extended functional roles*. As we have seen, social externalism is strictly speaking an active externalism. Speakers of a language are part and parcel of the language game and will in principle actively influence the overall usage of linguistic expressions. For natural languages this influence is of course negligible. This is why we may consider the socially extended functional roles as *effectively extended long-arm roles*. However, for the sake of accuracy and in order not to confuse this terminology with the idea of long-arm roles of natural-kind and historical externalism (the latter are sometimes dubbed 'adaptational functional roles', cf. Cummins 1989), I rather propose to distinguish between *degrees* of functional role extension. In contrast to extended functional roles that comprise vehicles of the closely coupled proximal environment and are in this sense of short-range extension only, language-use roles generally are of long-range extension (that was already indicated in 3.1 by considering "functional roles of our language use that are already part of the socially extended system"). How far they really reach is a matter of degree and scales with the size of the language community (in tune with our result from 3.3). Notice that the short-range/long-range distinction is a distinction of *degrees of active extension*, while the short-arm/long-arm distinction is categorical, insofar as it pertains to the distinction between internalism and passive externalism.

As a version of FRS, an EC inspired semantics must of course come to terms with the

known problems of FRS, as, for instance, worries about semantic holism and the alleged inability to account for a compositional semantics. Again, an analysis of these questions goes beyond the present paper, but supporters of FRS have countered such objections (cf. Block 1998 and Whiting 2010). It is therefore unclear how serious these objections are to begin with. Probably, as far as I can see, an extended cognition perspective on FRS adds nothing new, either in positive or negative terms, to these debates.

Truly positive prospects might lie elsewhere. We have seen that extended functional roles go beyond the originally conceived short-arm functional roles. The extended system is already an intrinsically world-bound and environment-coupled system. This genuine situatedness of extended cognitive systems offers a fascinating perspective on the notorious problem of *semantic grounding*. The traditional dichotomy between the representational system and the target of its representation or, perhaps more generally, the subject/object divide, gets blurred. Parts of the environment can be promoted to become extended vehicles and, thus, to become parts of the cognitive system. This means that some of the causal connections of (the components of) the vehicles play extended functional roles. At the same time, the vehicles are causally connected to the greater environment in which the extended system is embedded. Thus, causal connections that play extended functional roles and their prolonged, ordinary causal connections are more intertwined than traditionally conceived. I would like to suggest that in a properly worked out theoretical setting the semantic grounding of cognitive systems can perhaps be traced back to this peculiar feature of extended functional roles – without invoking mysterious passive externalist long-arm (or extended long-arm) functional roles.

This suggestion and many of my other remarks on extended functional roles are in need of a more rigorous elaboration within a future 'extended functional-role semantics'. But we can nevertheless see the merits of the foregoing discussion. Uncovering social externalism as a sub-variant of active externalism helps us to get a first grip on the intriguing question of whether there exists a specific theory of meaning connected with active externalism. Since the context of discovery for natural kind, historical and social externalism was decidedly different than that for active externalism, no such connection has hitherto been spelled out. But the above analysis provides us with a rough sketch of such an EC inspired semantical account: an extended FRS account that is in tune with active externalism and that embraces a use theory of meaning via active social externalism.

Acknowledgement.

The present version of the paper owes much to the kind suggestions of Kristina Musholt, Michael Pohl and an anonymous referee of this journal.

References

Adams, F. R. & K. Aizawa 2010: Defending the Bounds of Cognition. In *Menary (2010)*, 67-80.

Burge, T. 1979: Individualism and the Mental. *Midwest Studies in Philosophy*, 4, 73-121

Block, N. 1998: Semantics, Conceptual Role. In: E. Craig (ed.), *The Routledge Encylopedia of Philosophy*. London: Routledge.

Blomberg, O. 2011: Socially Extended Intentions-in-Action. *Review of Philosophy and Psychology*, 2(2), 335-353.

Chalmers, D. 2008: Foreword. In *Clark (2008)*.

Chalmers, D. (ed.) 2002: *Philosophy of Mind: Classical and Contemporary Readings.* Oxford: Oxford University Press.

Child, W. 2006: Wittgenstein's Externalism: Context, Self-Knowledge and the Past. In: T. Marvan (ed.), *What Determines Content? The Internalism/Externalism Dispute*. Newcastle: Cambridge Scholars Press.

Clark, A. 2005: Intrinsic Content, Active Memory and the Extended Mind. *Analysis*, 65(285), 1-11.

Clark, A. 2008: *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Clark, A. & D. Chalmers 1998: The Extended Mind. *Analysis*, 58(1), 7-19.

Cummins, R. C. 1989: *Meaning and Mental Representation*. Cambridge, MA: MIT Press.

Hurley, S. 1998: Vehicles, Contents, Conceptual Structure, and Externalism. *Analysis*, 58(1), 1-6.

Hurley, S. 2010: Varieties of Externalism. In *Menary (2010)*, 101-153.

Jacob, P. 1993: Externalism and the explanatory relevance of broad content. *Mind and Language* 8(1), 131-156.

Keijzer, F. & M. Schouten 2007: Embedded cognition and mental causation: Setting empirical bounds on metaphysics. *Synthese* 158, 109-125.

Kirsh, D. & P. Maglio 1994: On distinguishing epistemic from pragmatic action. *Cognitive Science* 18: 513–49.

Kosslyn, S. M. 2006: On the Evolution of Human Motivation: The Role of Social Prosthetic Systems. In S. M. Platek, T. K. Shackelford & J. P. Keenan (eds.): *Evolutionary Cognitive Neuroscience*. Cambridge, MA: MIT Press, p. 541-554.

Krueger, J. 2011: Extended cognition and the space of social interaction. *Consciousness and Cognition*, 20(3), 643-657.

Lyre, H. 2009: The "Multirealization" of Multiple Realizability. In: A. Hieke and H. Leitgeb (eds.): Reduction - Abstraction - Analysis. Ontos, Frankfurt, p.79-94.

Lyre, H. 2010: Erweiterte Kognition und mentaler Externalismus. *Zeitschrift für philosophische Forschung* 64(2): 190-215.

Lyre, H. forthcoming: Socially Extended Cognition and Shared Intentionality. Preprint.

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

Robbins, P. & M. Aydede (eds.) 2008: *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.

Rowlands, M. 2003: *Externalism: Putting Mind and World Back Together Again*. Chesham: Acumen.

Whiting, D. 2010: Conceptual Role Semantics. *Internet Encyclopedia of Philosophy*.

Wilson, R. A. 2010: Meaning Making and the Mind of the Externalist. In *Menary (2010)*, 167-187.

Wittgenstein, L. 1953: *Philosophical Investigations*. Oxford: Blackwell.