Manuskript

Titel: **“Seeing oneself through the eyes of others. Beckermann on self-consciousness”**

Autoren:

Hofmann, Frank, Prof. (Luxemburg)

Pöhlmann, Ferdinand, M.A. (Tübingen)

Korrespondenz:

Prof. Frank Hofmann
Universität Luxemburg
Department of philosophy
FSLHASE, IPSE
Campus Walferdange
Route de Diekirch, B.P. 2
L – 7220 Walferdange
Luxemburg

frank.hofmann@uni.lu
Tel.: 00352 466644 6639 (Handy: 0049 24546778)

“Seeing oneself through the eyes of others.
Beckermann on self-consciousness”

Abstract

Ansgar Beckermanns Theorie zur Erklärung von Selbstbewusstsein kann als exemplarischer Versuch verstanden werden, die Wurzeln selbstbewusster Zustände in sozialer Kognition zu suchen. Dabei wird angenommen, dass eine notwendige Bedingung zur Entwicklung von Selbstbewusstsein darin besteht, dass sich kognitive Wesen in sozialer Interaktion ‚mit den Augen eines anderen sehen‘ lernen. Dieser Ansatz scheitert aber aus prinzipiellen Gründen, da er die besondere, identifikationsfreie Referenz von Ich-Gedanken nicht erklären kann und damit eine Erklärung wesentlicher Züge von *de-se*-Zuständen schuldig bleibt. Zudem ist Beckermanns Theorie im Besonderen durch einige weitere, grundlegende Unzulänglichkeiten gekennzeichnet.

Ansgar Beckermann’s account of self-consciousness can be seen as an attempt to locate the origin of self-conscious states in social cognition. It is assumed that in order to acquire self-consciousness, a cognitive system has to ‘see itself through the eyes of the others’. This account, however, is doomed to failure, for principled reasons. It cannot provide a satisfactory explanation of the special, identification-free reference of first-person thoughts and, thus, fails to explain crucial features of *de-se* attitudes. In addition, Beckermann’s account exhibits various other shortcomings.

"Seeing oneself through the eyes of others.
Beckermann on self-consciousness"

1. Introduction

Our primary goal in this paper is to describe and criticize a certain approach to self-consciousness which takes self-consciousness to originate in social cognition. As a paradigmatic and particularly clear example of this approach we take Ansgar Beckermann's specific account of self-consciousness (Beckermann 2008 and Beckermann 2003). According to Beckermann, the ability to think of oneself as oneself, constitutive of self-consciousness, arises from the task of representing the contents of others' mental states – as (sometimes) concerning oneself.

The first part of our paper summarizes Beckermann's account of self-consciousness (section 2). The rest of the paper is dedicated to a critique of this account. Some not too serious problems with the proposal are briefly mentioned and put to one side (section 3). These problems concern certain aspects of Beckermann's proposal which are of a quite general nature, like the notions of computational and causal role and the conceptual and indexical character of first-person states. Even if Beckermann could deal successfully with these problems, however, serious problems concerning directly the nature of self-consciousness would remain. These problems constitute the main difficulty for Beckermann's account, and we will discuss them in detail (section 4). The main difficulty is a generic problem for any 'social' account of self-consciousness: it cannot explain the peculiar *de-se* character of self-conscious states, since it cannot explain the peculiar reference of the first-person concept that is immune to error through misidentification.

2. Beckermann's account of *de-se* states

Let us begin with an overview of Beckermann's theory of self-consciousness which can count as a paradigmatic example of the approach towards an explanation of self-consciousness via

social cognition.[1] To illustrate his account, Beckermann tells a story of a cognitive system, called 'Al'. Although one could get the impression that Beckermann takes this story to be a real account of the ontogenetic development of self-representations and self-consciousness, we think one should take it merely as a metaphor or illustration which is meant to describe various forms of self-knowledge and their natures, functions, and limitations. Accordingly, we will describe his story not as a series of developmental steps of one organism, but we will carve out and present the central claims of the view standing behind the story.[2]

Beckermann starts by noting that only entities that form *representations* of their environment are capable of having self-representations. This means that simpler organisms that admittedly can experience their environment and behave in it but do not form representations,[3] are *per se* not capable of forming self-representations. An example would be a system whose behavior rests exclusively on stimulus-response-mechanisms.[4]

Organisms that form representations of their environment are termed 'cognitive systems' by Beckermann. Drawing on John Perry's work,[5] he claims that cognitive systems represent objects in their vicinity as standing in certain spatial and other relations to themselves due to the agent-relative roles these objects play with regard to them.[6] This means that cognitive systems do not represent the objects in their environment in some kind of universal coordinate system as, e.g., the AI-program SHRDLU does. Instead, they represent objects in an egocen-

[1] Another example for this kind of approach is Musholt (2012). Tugendhat (1979) is sometimes interpreted as belonging to this approach, but he has rejected this interpretation (Tugendhat 2005). Of course, the approach goes back to, or is inspired by, George Herbert Mead's philosophy, and it is manifest in Jürgen Habermas' work, in one way or another.

[2] To give just one example of a question that would have to be addressed in a genuine, serious account of the ontogenetic development of self-consciousness: How can an organism that has absolutely no conception of itself come to represent that another subject is looking at *it*, or wants to interact with *it*? At least some rudimentary form of self-representation seems to be required. In general, Beckermann's story is rather a theory of the nature and function of self-consciousness than a developmental theory.

[3] Beckermann identifies representations as "more or less stable internal structures, that stand for some aspects of the environment, even if they are not currently experienced." Beckermann (2008, 69). (Our translation.)

[4] Beckermann gives as examples the crab that Churchland uses to illustrate eye-arm-coordination in Churchland (1986, 284), and the program SHRDLU, designed by Terry Winograd. Cf. Beckermann (2008, 68) and Beckermann (2003, 176), respectively.

[5] Especially Perry (1998).

[6] In Perry's terms, agent-relative roles of objects determine epistemic and pragmatic methods that are appropriate to get to know of or act on them, respectively. See Perry (1998, 84), and Perry (2002, 197ff.) (Here, his terminology has changed from 'agent-relative *roles'* to 'agent-relative *relations'*.)

tric frame of reference. One might think that this mode of representation presupposes some kind of self-representation, but Beckermann explicitly denies this claim by stating that in order to egocentrically represent objects around oneself one merely has to represent the property of standing-in-a-certain-relation-to-oneself. No explicit self-representation is contained.[7] The relevant properties are agent-relative properties, and one can represent them '*en bloc*', as it were. The same holds true of the case of representations of the cognitive system itself: it is possible to represent internal states of the representing system without explicitly representing the system as such. According to Beckermann (and Perry, too), the reason for this is simply the fact that the kinds of representations that figured hitherto in the discussion – egocentric/agent-relative representations of the environment and proprioceptive representations – share the property that every token of them always concerns one and the same cognitive system, namely, the one that these representations belong to. It will never be the case that a particular cognitive system represents objects in its environment as standing in a relation to a cognitive system other than itself when engaged in egocentric perception; equally, it will never be the case that a cognitive system represents internal states of some other cognitive system when engaged in proprioception. Nevertheless, such representations amount to a kind of self-knowledge in that they implicitly purport to carry information about the spatial relations the representing system bears to the represented objects and about the internal states that are actually instantiated in the representing system. For that reason, Perry calls such knowledge 'agent-relative knowledge'.

By now, we have an organism which is capable of perceiving, and behaving in, its environment by means of its agent-relative knowledge. No explicit self-representation – i.e., no representation of the representing system itself – is needed for these purposes. But the world does not only contain dead objects but *other cognitive systems* as well. Hence, it is very important for each cognitive system to know what the others are going to do, and whether they are friendly or hostile against oneself, in order to act appropriately. This is why cognitive systems not only represent objects and their properties but other cognitive systems and their mental states as well. In other words, cognitive systems are capable of *meta-representations*, i.e., representations of representations. The next step is a crucial one: in order to represent representations of others which have as their content the representing system itself, the representing system has to represent itself. That's where a self-representation comes into play. This

[7] This claim, although not questioned by Beckermann, is not entirely uncontroversial. For there are accounts that presume some form of self-representation even in egocentric representations of the environment, see e.g. Schellenberg (2007) and, less clearly though, Brewer (1992).

self-representation is the core representation of a fully self-conscious system. Henceforth, it is not only used to represent others' representations of oneself, but also to represent one's own physical and mental properties and to develop a body schema.[8]

But what exactly makes this self-representation a *self-conscious* representation of oneself, i.e., a *de se*-representation? All we have so far is a representation of a particular cognitive system that happens to be of the cognitive system itself. Now Beckermann claims that exactly those self-representations count as *de se* which are *equivalent* to the agent-relative representations the organism uses to orient in and interact with its environment. Equivalence in this case means *having the same computational role*.[9] This has two decisive consequences. Firstly, proprioceptive input leads not only to agent-relative representations but to explicit self-representations as well. Secondly, explicit self-representations inherit two futures of agent-relative representations concerning actions. The first feature is that in particular situations, representations with agent-relative content immediately lead to certain actions, typically. Representations that contain explicit reference to the representing system are then supposed to have the same immediate impact on action. For example, when a ball is flying in one's direction, one will immediately try to catch it or to avoid being hit by it, typically. This will be the case no matter whether one represents merely the agent-relative roles of the ball or whether one explicitly engages in self-conscious representation.

The second feature is that both types of representations are connected with certain bodily movements that constitute certain types of actions. That is, to represent in an agent-relative manner some object of one's vicinity means directly to know how to move one's body in order to interact with that object. This feature is also supposed to be transferred to the corresponding self-conscious representations.

With these two features inherited from agent-relative representations, self-representations play a special role among all representations of a cognitive system, according to Beckermann. The special role lies exactly in the action-related consequences those self-conscious representations have. As an illustration Beckermann cites Perry's famous example of a supermarket shopper who follows a sugar trail, trying to tell the owner of the damaged sugar package that he is making a mess and not realizing that he himself is the one with the damaged package. When he finally realizes his fault, his actions concerning 'the owner of the damaged sugar package' change significantly.[10]

[8] Beckermann (2003, 183f.).

[9] Beckermann (2003, 186).

[10] Perry (1993, 33).

To sum up Beckermann's account, one can say that *de-se* representations have the following distinguishing marks: they are used to represent others' representations of oneself and one's own physical and mental states, and they have an immediate significance for action due to their equivalence with agent-relative representations. In Beckermann's words: "*[D]e se* beliefs […] are characterized by the specific causal role they play within the cognitive system of a person and with regard to her or his actions."[11]

3. Problems with Beckermann's account

In this section we will briefly note several problems for Beckermann's account. These concern certain aspects that are usually associated with self-conscious states in the philosophical debate. We will mention four problems: the issue of indexicality, a problem concerning causal-computational role and Beckermann's thesis of 'equivalence', the issue of the conceptual character of self-consciousness, and the issue of the identity of the first-person concept (over time). We think that these four issues raise serious questions and problems, but we put them to one side. The main purpose of this section is to prepare the ground for presenting the real, fundamental difficulty, by setting it apart from the four issues just mentioned. The fundamental difficulty with Beckermann's account (and, indeed, any 'social' account of self-consciousness) will be presented in the next section.

A first question concerns the *indexicality* of the first-person concept. Beckermann relies heavily on ideas of John Perry concerning different types of 'self-related knowledge.' He incorporates in his story the notions of knowledge of the person one happens to be, agent-relative knowledge and self-attached knowledge, the fundamental triad in Perry's work. But regarding another central claim of Perry's, namely the essential indexicality of the first-person concept, Beckermann does not say a word. Perry claims that some indexical expressions, namely, those that express 'locating beliefs' about who one is, which time it is, and where one is, are not substitutable by other (non-indexical) expressions without loss of explanatory force.[12] It's not clear whether Beckermann regards this feature of 'I' as non-essential as, e.g., Ruth Millikan does,[13] or whether he thinks that in his tale about the genesis of self-representations the indexicality is somehow incorporated into his account. The first interpreta-

[11] Beckermann (2003, 186).

[12] Perry (1993).

[13] Millikan (1990).

tion is supported by the fact that he claims that all objects, including the representing system itself, are represented by internal 'names,' which by definition are not indexical expressions, we take it. But the second interpretation seems plausible as well, since Beckermann claims that Al would express representations that are about itself by the indexical 'I.'[14]

A second issue concerns causal-computational role and Beckermann's 'equivalence thesis'. A central element in Beckermann's account is the notion of a 'role' a representation has. On the one hand, there is the special *computational role* that the self-representation has.[15] On the other hand, there is its special *causal role*.[16] The causal role consists in the special, unmediated action-relevance that agent-relative representations have and that is inherited by the self-representation via its equivalence with agent-relative representations. This sounds quite right, and it is widely accepted that *de se*-representations have this kind of action-relevance. But the problem with Beckermann's account regarding this point is twofold. Firstly, he gives no explanation of the unmediated action-relevance of agent-relative representations, he only states this supposed fact by re-telling some examples from Perry. These examples in turn rely on the unargued assumption that Fregean modes of presentation can be interpreted as causal roles. Even if this is the case, neither Beckermann nor Perry provides any explanation of it. The second problem lies in the postulated equivalence between agent-relative representations and self-representations. Even if we take it for granted that agent-relative representations have this special causal role, Beckermann does not put forward anything which could make it intelligible that self-representations inherit this role – he merely postulates this. To give some evidence how this might happen, Beckermann should have presented (within his story) a somewhat more detailed account of how the equivalence between agent-relative representations and self-representations evolves. But he only states that it does evolve.

It seems, though, that the *causal role* is dependent on the *computational role* a representation has in the cognitive architecture of a cognitive system. Beckermann states that the causal role is a 'consequence' of the equivalence between agent-relative and self-representations, and he writes that "'being equivalent' here mean[s] simply 'having the same computational role'".[17]

So, perhaps we just have to look at the computational roles of these representations. Unfortunately, the situation here is the same, if not worse. The computational role is explained

[14] Beckermann (2003, 184).

[15] Beckermann (2003, 186); Beckermann (2008, 77f.).

[16] Beckermann (2003, 186f.); Beckermann (2008, 79ff.).

[17] Beckermann (2003, 186).

by Beckermann in terms of 'modes of presentation' or Fregean contents: representations with the same mode of presentation/Fregean content correspond to the same computational role.[18] Hence, self-representations have a special mode of presentation, which Beckermann calls "EGO-mode of presentation".[19] But Beckermann does not tell us more about it, except that it is a "very special way"[20] in which one is given to oneself. Furthermore, Beckermann states that the computational role of self-representations has *two* special features. The second feature is the peculiar *causal role* we already encountered. The first feature is that the proprioceptive input is not only represented in terms of agent-relative representations but also explicitly in terms of self-representations. Hence, it seems that the computational role of self-representations is simply special because proprioceptive input is related to them and they have a close connection to action. But if we take Beckermann literally, even these features are not peculiar to self-representations, since agent-relative representations have exactly the same properties. For, the equivalence between the two ensures that they have the same computational role.

A third problem with Beckermann's account concerns the *conceptual character* of self-representation. Beckermann writes that only "self-knowledge in a strong sense" requires that the cognitive system develops a concept of itself.[21] This seems to imply that the self-representation is a concept. But Beckermann does not say anything about the consequences that being a concept could have for the self-representation. Instead, he claims that conceptual self-representations and agent-relative representations, which form a kind of self-knowledge that is not dependent on a concept of oneself, are very similar in that they have the same computational roles.[22] Does this mean that there exist non-conceptual and conceptual forms of self-consciousness side by side? Or is only the second kind a kind of self-consciousness? Does the difference between the two only consist in that the second one refers explicitly to the representing system while the first one does so only tacitly or implicitly? If that is the only difference, wouldn't that mean that conceptual and non-conceptual representations of particular things are only different in their degree of explicitness? Beckermann does not say anything that could give us an answer to these questions.

[18] Beckermann (2003, 185).

[19] Beckermann (2003, 185).

[20] Beckermann (2003, 185).

[21] Beckermann (2008, 75).

[22] Cf. Beckermann (2008, 75).

Another point relating to the issue of conceptual character is the question of the *identity* of the self-representation. Beckermann claims that the first-person concept develops from the social interaction and the need for meta-representations of others' mental states coming with it. Hence, at the beginning we have a self-representation that has a specific role among the representations of a cognitive system, namely, to represent oneself as one figures in mental states of others. But later, it will have the computational role mentioned above. It seems that important features of the first-person concept have changed, and one could wonder whether the first-person concept before the development is the same concept as the concept after it. Because one could argue that concepts are (partly) defined by their computational role, it seems at least doubtful that this is the case. Unless Beckermann gives an argument in favor of the identity of the relevant concept over time, the story might seem incoherent.[23]

4. The fundamental difficulty with Beckermann's theory

Aside from the problems just mentioned we think that the fundamental difficulty with Beckermann's account – and with the entire 'social approach' to self-consciousness – resides in the fact that *it cannot explain the peculiar* de-se *character of self-conscious representations*. The means available to Beckermann are not suitable for the task of explaining the *de-se* character of self-representations. (This problem, we submit, is generic to the entire 'social approach' to self-consciousness, and not just a problem for Beckermann's specific version.) One cannot explain the *de-se* character of self-conscious states by reference to a role in social metacognition. This is what we would like to argue in the remainder of this paper.

The fundamental difficulty relates to two important features of self-consciousness: its *reference* and its *immunity*. Let us explain. The first-person concept refers, we take it, *pace* Anscombe. It is self-referential, in the sense of referring to the subject. For example, Al's first-person concept refers to Al. (Self-referentiality in this sense is simply reflexivity, and does not require or involve any *de-se* mode of presentation.) But the first-person concept refers in a special way. It refers in a way which allows for representations which are immune to

[23] Originalism about concept identity, as recently proposed by Sainsbury, Tye (2011), would be a view of concepts that is favorable to Beckermann's theory. Originalism allows for changes in semantic features and/or computational roles, without loss of identity of the concept. Other views of concepts that individuate concepts semantically (such as Fodor's representational/computational theory) seem to be incompatible with semantic changes.

error through misidentification with respect to the first person. For the sake of brevity, let us call this feature 'immunity'. Immunity in this sense lies at the heart of the special *de-se* character of self-conscious states. Indeed, one can take it as the crucial criterion for self-consciousness. The debate about self-consciousness has focused on immunity ever since Strawsons's, Shoemaker's, and Evans' works. We will follow their lead here. So let us suppose that what needs to be explained about the first-person concept is its reference and immunity.[24]

The crucial question now is whether Beckermann's account can explain the reference and immunity of the first-person concept. In the following we would like to argue that the answer is negative. Beckermann's account lacks the resources for such an explanation. We take this as a sufficient reason for concluding that his account cannot explain self-consciousness, since reference and immunity are the crucial features of self-consciousness in need of explanation.

Let us begin by taking a closer look at *immunity*. Roughly, a first-person representation is immune (with respect to the first-person position)[25] just in case it is impossible that it is false because, and only because, the subject represents of someone that something is true of that person, but fails to identify that person with herself.[26] For example, the first-person representation whose content can be expressed by the utterance 'I see a pink elephant' can be false because the elephant I am currently seeing is in fact grey, but it cannot be false because it is in fact you who sees the elephant but not me (at least, normally). Hence, this representation is immune to an error on the ground that I misidentify someone seeing an elephant with me. Whether the representation is immune or not is not determined by the object the judgment refers to, but by the kind of information that grounds the judgment (that is, the way in which the judgment is arrived at).[27] The best characterization of immunity is given by *identification-freedom*. Because I do not identify myself, there is no possibility of mis-identification. This is why I cannot misidentify someone else as me. Immune first-person representations are identi-

[24] Bermúdez (2011) takes immunity (in this sense) as definitive of self-consciousness. We take it that immunity is a necessary condition, not a sufficient condition.

[25] Strictly speaking, immunity is always relative to a certain position. For the sake of brevity, we will not always mention this qualification in the following. The context should make things clear enough.

[26] Shoemaker develops this property with reference to Wittgenstein in Shoemaker (1968).

[27] Cf. Gertler (2011, 216); Evans (1982, 218f.). With regard to the first-person concept this characterization implies that not every judgment involving it is immune to such an error. And in general, other judgments concerning particular things (*de re*-judgments) can be immune in that sense, too. See Gertler (2011, 216) and Evans (1982, 219) for examples.

fication-free. Reference to myself is not mediated by any detection of identifying properties, i.e., it is not mediated by any identification of myself.[28]

Immunity can be spelled out in detail in various different ways. But what we will argue will not depend on which of these various more precise statements of immunity one favors. The crucial feature of immunity, for our purposes, is identification-freedom. This is, more or less, Evans' understanding of the phenomenon.[29] And from now on, we will rely on this basic understanding. Our argument will be independent of any further details.[30]

Let us now consider the issue of *reference*, and the explanation of reference, of the first-person concept. For many philosophers, the first-person concept is referential, i.e., each token refers to a certain individual (namely, the one which is exercising the first-person concept on that occasion). Famously, Elisabeth Anscombe has denied that the first-person concept refers.[31] But it seems fair to say that her view is rather an implausible position, and there are not many who have followed her. Now, let us suppose that the first-person concept is a referential representation. (And clearly, Beckermann agrees.) This raises the question how we can explain its reference. Following Reichenbach, we can say that the token-indexical rule *describes* the reference of the first-person concept: any token of the first-person concept, occurring within the thought t, refers to the thinker of this thought t. This, however, does not provide an *explanation* of why a token of the first-person concept refers to the subject to which it in fact refers. And one can wonder whether it is not an important theoretical task to illuminate and explain how the first-person concept refers.

---

[28] Here we want to point out that we distinguish between referring to oneself and identifying oneself. So referring to oneself is not *per se* an identification of oneself. Identification requires the detection of some property, or cluster of properties, sufficiently rich for determining the referent. Reference to oneself is possible without the detection of such a property, or cluster of properties. This is one of the lessons we take from Shoemaker. Note, furthermore, that the property, or cluster of properties, used for identification can contain indexical-demonstrative elements. (Récanati, 2007, uses 'identification' and 'articulation' more or less synonymously. Cf. ibid., 147, e.g. In this sense, then, reference is also to be distinguished from articulation.)

[29] Cf. Evans (1982, 180f.).

[30] Perhaps, one should distinguish between 'absolute' and 'circumstantial immunity', as Shoemaker (1968) does. (Récanati thinks that Shoemaker is confused here. Cf. Récanati, 2007, ch. 20.) And perhaps, immunity can be extended to the predicative position, as Bar-On (2004) suggest. Since we are interested only in the first-person concept, we will ignore any such extension. For further refinements and discussions of immunity see, for example, Pryor (1999), Coliva (2006), Récanati (2007). For our purposes, the basic understanding of immunity as identification-free reference and self-knowledge is sufficient.

[31] Cf. Anscombe (1975).

If we take together these two features of the first-person concept, its reference and its immunity, it seems quite clear that in principle, an explanation of reference could be given by recourse to the causal-computational role – and, indeed, an explanation which is perfectly in line with immunity. The basic idea for such an explanation is the special causal-computational role that the first-person concept has in relation to egocentric perception and proprioception. The first-person concept is tied *immediately* (i.e., without any mediation by identifying properties) to proprioceptive experiences that represent the subject's bodily states. If, for example, Al feels pain in his left knee, Al is inclined to immediately form the first-person representation 'Al feels pain in his left knee.' (This is so if Al's internal name 'Al' really is a first-person representation.) Similarly, 'Al' is immediately linked to egocentric contents in Al's perceptual experience. If Al's perceptual state represents a tree-in-front-of-Al, then Al is inclined to immediately form the first-person representation 'A tree is in front of Al.' Proprioception and egocentric perception provide information about Al to which Al's first-person concept is sensitive *without the help of any mediating identification.* Now, plausibly, because the information is always about Al, this concept refers to Al. As Beckermann emphasizes, proprioception and egocentric perception always provide information about one and the same object, namely, the cognitive system itself (Al). And this is why no identification is required. If – perhaps *per impossibile* – proprioception could sometimes concern some other cognitive system, identification would become necessary. Only because there is no such variation in the object of proprioception, identification is superfluous. We have a kind of (structural, non-accidental) 'informational constancy' which makes identification superfluous. (Similarly, egocentric content always relates things to one and the same cognitive system, Al.) As Récanati nicely puts it: "The subject himself does not need to be explicitly represented, since the representation can only be about him and his situation."[32] The concept that Al acquires on the basis of proprioception and egocentric perception allows for immune self-representations, since it refers to Al and does so without any identification of Al (i.e., without the need for any uniquely identifying descriptive content).

We can generalize the result. Whenever a system has a set of representational states that always concern the system itself – which exhibits informational constancy –, it seems possible to 'introduce' a concept of the system which allows for immune self-representation. By its causal-computational role, this concept is sensitive to these representational states, and it can refer without identification.

[32] Récanati (2007, 146). Here, Récanati echoes Perry's talk of 'implicit' and 'explicit representation'. Perry also speaks of the subject's being an 'unarticulated constituent' of the representation. Cf. Perry (1986).

By now, surely we do not yet have a full-blown explanation of reference and immunity. But at least, we have an idea and a sketch of such an explanation. And it does not seem hopeless to think that the searched-for explanation could be given along these lines. So we have reason to assume that this is the right track for explaining reference and immunity.

This raises a problem for Beckermann's account. According to Beckermann, the origin of Al's first-person concept is social metacognition. The primary job of the first-person concept is to represent Al as occurring in the contents of others' representations. Now, it may be the case that the first-person concept performs this job. But does this help to explain its crucial features, reference and immunity? On reflection, the answer is negative. It seems hopeless to try to understand reference and immunity by looking at the role of the first-person concept in social metacognition. The reason for this is simply the fact that *there is no informational constancy* to be found here – in contrast to the just-mentioned explanation in terms of proprioception and egocentric perception. Others do not always represent Al, they represent other conspecifics as well. Probably, they will represent Al only in a minor fraction of all cases. So there is not even any approximation of informational constancy. Al has to find out whether another cognitive system represents Al or some other, third cognitive system. It would be wildly incorrect to assume (by default) that others always represent Al. Therefore, the explanatory idea just mentioned cannot be transferred to, or mirrored within, the social metacognition account. Something else needs to be provided as an explanation of reference and immunity, and it is hard to see what could do the job. At least, Beckermann does not provide it, and we cannot even see any hint in his account.

Beckermann holds that Al's internal name of Al comes to acquire a certain causal-computational role (the one he tries to describe by speaking of 'equivalence'). The alleged 'equivalence' between Al's internal name and states with agent-relative content consists essentially in a certain causal-computational role of this name, and it is an important element in his account. At the same time, Al's internal name of Al is used, by Al, in order to represent Al as occurring in the representational contents of others' states. So the internal name plays two important roles at the same time. Now, however, the question is which role explains what. And given the difference with respect to informational constancy just pointed out, it seems clear that the situation is quite asymmetric. The causal-computational role explains reference and immunity, the role in social metacognition does not. If this is so, we have to conclude that the origin of the first-person representation lies in the system's own representational states exhibiting informational constancy (proprioception and egocentric perception), not in social metacognition.

The following diagnosis seems plausible. It is not an accident that the connection to proprioception and egocentric perception occurs in Beckermann's theory. This part is doing the explanatory work, for the explanation of reference and immunity. Once we have that work done, the first-person concept can be recruited for some other task, such as the task of social metacognition and theory of mind. But Beckermann commits a mistake if he locates the origin of self-consciousness in social metacognition. It is another and distinct part of his overall theory which explains self-consciousness – or, at least, could provide the basic material for such an explanatory account. The origin of self-consciousness lies in whatever accounts for reference and immunity, if it lies anywhere at all. Al can begin "to see himself through the eyes of others", but only if, and since, Al already possesses self-consciousness.[33]

Finally, let us take a look at a possible argument in defense of Beckermann. Beckermann might suggest that his goal was to explain self-consciousness in so far as it is necessary for dealing with a certain task; for other tasks, representations with agent-relative contents are sufficient. Therefore, self-consciousness has its 'origin' in social metacognition. Or so Beckermann might claim.

This argument fails, however, and it fails for two reasons. First of all, what is claimed within this argument is not correct, viz., that a self-conscious representation of Al is only necessary when it comes to social metacognition. And secondly, even if this were correct, it would not improve the situation with respect to the question of how we can explain the *de-se* character of the first-person concept. An explanation of immunity and reference would still be lacking. We would still not understand how Al's representation of Al could be self-conscious. Let us argue for these two points in the following.

Firstly, the argument in defense of Beckermann contains a false claim. The claim is that a self-conscious representation of Al is only necessary when it comes to social metacognition. To see why this claim is false, one has to consider the issue of the *form* or *format* of representation, i.e., of the distinction between conceptual and non-conceptual content. Arguably, proprioception and egocentric perception have non-conceptual content, whereas thoughts and other propositional attitudes possess conceptual content. We take it as an empirically well-established fact about human cognition that perception (including proprioception and imagery) is processed in a way which is quite different from the way in which conceptual categorizations are processed. The best explanation for this is the assumption that they differ in the form or format of representation. If this is so, the need for acquiring a first-person *concept* is al-

[33] Beckermann (2003, 184).

ready in place when it comes to information about the system's own bodily state (proprioception) and the agent-relative roles of objects in the system's environment (egocentric perception). The proprioceptive and perceptual states are not suitable for thought, since they do not have the requisite kind of form or character – they are not concept involving. Therefore, if one accepts a distinction between conceptual and non-conceptual representation and assigns non-conceptual content to proprioception and egocentric perception, Beckermann's argument fails. Even if all the information is already contained in these non-conceptual states, a first-person concept is needed in order to bring it into the realm of thinking (with all its inferential capacities and processing). Al not only wants to (proprioceptively) *perceive* the pain in his left knee. Al also wants to be able to *think* that there is a pain in his left knee. Therefore, Al needs a first-person concept. Without such a concept Al's thinking would be 'blind' to the information contained in his perceptual states.[34]

Beckermann could resist this counter-argument by rejecting the distinction between conceptual and non-conceptual representation. (We believe that this would be a rather desperate move.) But even then his argument would not succeed. For, consider egocentric contents in perception. These perceptual states represent agent-relative properties of objects in Al's environment. For example, a red apple is represented as red and as being-one-meter-in-front-of-Al. The spatial feature of the apple is represented *en bloc*, as it were. But certainly, Al not only wants to represent apples as having spatial locations relative to Al. He also wants – and needs – to represent apples as standing in the very same spatial relations to other objects. Therefore, Al needs a representation of the spatial relation being-one-meter-apart-from, and not only of the impure spatial property of being-one-meter-in-front-of-Al. And then he needs an explicit representation of Al in order to apply the former representation of the spatial relation, if he wants to represent Al as standing in this relation to some other object. Of course, whether Al 'needs' a certain representation depends on the tasks Al is supposed to solve. So 'needing' a kind of representation is relative. But it seems quite clear that the 'need' for a representation of spatial relations (and not just of impure, agent-relative spatial properties) is quite strong, since it allows for a *much more general* application which is useful for a potentially unlimited number of cases. (Of course, many inferential relations will become accessi-

[34] The distinction between conceptual and nonconceptual content has by now become very wide-spread. It can be drawn and explained in various different ways, e.g., by reference to maplike vs. sentence-like representation (Tye) or analog vs. digital representation (Dretske), etc. It does not matter for the present purposes which of these further developments or explanations one favors, all we need is the distinction itself. It does also not matter whether the distinction concerns the contents or the representations (representational vehicles) or both.

ble only by having the complex representation, comprising two singular representations of the objects and a representation of the spatial relation. Only then will many logical relations become 'visible'.)[35]

If, on the other hand, Beckermann wanted to insist that the 'need' for a self-representation that comes from the task(s) of social metacognition is strict and absolute, we have to respond that, on reflection, this is not really true. In principle, one could represent others' mental states concerning oneself in an *en-bloc* fashion, in the same way in which egocentric representations represent the spatial facts in an *en-bloc* fashion. For example, Al would represent some other cognitive system as having the property of representing-Al-as-friendly. Of course, such a way of representing social mental facts would be vastly impractical, perhaps to the point of being no longer computationally manageable. (The range of application would not be general, as one can put it.) But in principle such a way of representing these facts is possible. So even there the 'need' is not strict or absolute.

Now, let us move on to our second point. We submit that even if the claim that we have just criticized were correct, this would not improve the situation concerning the explanation of the *de-se* character of self-conscious representations. Suppose that Al needed an explicit representation of Al for the task of social metacognition. This would not show, however, how this representation gets the two crucial features of reference and immunity. It could still be the case that what explains these features is something quite different from the role in social metacognition; and in particular, it could still be the case that what explains these features is a causal-computational role vis à vis proprioception and egocentric perception. Therefore, we have to conclude that Beckermann does not explain self-consciousness by means of social metacognition – even if the claim that we argued against above were correct. The explanation of self-consciousness may still reside in something quite different from social metacognition.[36]

[35] Just to mention another 'need' for an explicit self-representation, the use of spatial cognitive maps seems to require an element suitable for marking the system's own position (real or imagined) within its map.

[36] Many thanks for valuable comments and points of criticism to Ansgar Beckermann, Hanspeter Mallot, Mark May, Wolfgang Röhrich, Peter Schulte, and Hong Yu Wong.

References

Anscombe, Gertrude Elizabeth Margaret, 1975: The first person. In: Guttenplan, Samuel D. (Hg.): *Mind and language*. Oxford: Clarendon Press, S. 45-64.

Bar-On, Dorit, 2004: *Speaking My Mind: Expression and Self-Knowledge*. Oxford: Clarendon Press.

Beckermann, Ansgar, 2003: Self-Consciousness in cognitive systems. In: Kanzian, Christian (Hg.): *Persons: An interdisciplinary approach: Proceedings of the 25th International Wittgenstein Symposium 11th to 17th August Kirchberg am Wechsel (Austria)*. Kirchberg am Wechsel: Österreichische Ludwig-Wittgenstein-Gesellschaft, S. 175-188.

Beckermann, Ansgar, 2008: *Gehirn, Ich, Freiheit: Neurowissenschaften und Menschenbild*. Paderborn: Mentis.

Bermúdez, José, 2011: Bodily awareness and self-consciousness. In: Gallagher, Shaun (Hg.): *The Oxford Handbook of the Self*. Oxford: Oxford University Press., S. 157-179.

Brewer, Bill, 1992: Self-location and agency. In: *Mind* 101, S. 17-34.

Churchland, Paul M., 1986: Some Reductive Strategies in Cognitive Neurobiology. In: *Mind* XCV, S. 279-309.

Coliva, Annalisa, 2006: Error through misidentification. Some varities. In: *The Journal of Philosophy* 103, S. 403-425.

Evans, Gareth, 1982: *The varieties of reference*. Edited by John Henry McDowell. Oxford, New York: Clarendon Press; Oxford University Press.

Gertler, Brie, 2011: *Self-knowledge*. New York: Routledge.

Millikan, Ruth Garrett, 1990: The Myth of the Essential Indexical. In: *Noûs* 24, S. 723-734.

Musholt, Kristina, 2012: Self-consciousness and intersubjectivity. In: *Grazer Philosophische Studien* 84, S. 75-101.

Perry, John, 1986: Thought without representation. In: *Proceedings of the Aristotelian Society* 137, S. 137-152.

Perry, John, 1993: The problem of the essential indexical. In: Perry, John (Hg.): *The problem of the essential indexical: And other essays*. New York: Oxford University Press, S. 33-52.

Perry, John, 1998: Myself and I. In: Stamm, Marcelo (Hg.): *Philosophie in synthetischer Absicht: Synthesis in mind*. Stuttgart: Klett-Cotta, S. 83-103.

Perry, John, 2002: *Identity, personal identity, and the self*. Indianapolis: Hackett.

Pryor, James, 1999: Immunity to error through misidentification. In: *Philosophical Topics* 26, S. 271-304.

Récanati, François, 2007: *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.

Sainsbury, R. Mark; Tye, Michael, 2011: An Originalist Theory of Concepts. In: *Aristotelian Society Supplementary Volume* 85, S. 101-124.

Schellenberg, Susanna, 2007: Action and Self-Location in Perception. In: *Mind* 116, S. 603-632.

Shoemaker, Sydney S., 1968: Self-Reference and Self-Awareness. In: *The Journal of Philosophy* 65, S. 555-567.

Tugendhat, Ernst, 1979: *Selbstbewusstsein und Selbstbestimmung*. Frankfurt/Main: Suhrkamp.

Tugendhat, Ernst, 2005: Über Selbstbewusstsein: Einige Missverständnisse. In: Grundmann, Thomas et al. (Hg.): *Anatomie der Subjektivität*. Frankfurt/Main: Suhrkamp, S. 247-254.