Forthcoming in: *Phenomenology and the Cognitive Sciences*

Handedness, Self-Models and Embodied Cognitive Content

Holger Lyre*

September 2007

Abstract. The paper presents and discusses the "which-is-which content of handedness", the meaning of left as left and right as right, as a possible candidate for the idea of a genuine embodied cognitive content. After showing that the Ozma barrier, the non-transferability of the meaning of left and right, provides a kind of proof of the non-descriptive, indexical nature of the which-is-which content of handedness, arguments are presented which suggest that the classical representationalist account of cognition faces a perplexing problem of underdetermination of reference of left and right in the which-is-which sense. By way of contrast, no such problems occur in a framework were embodied contents are not mediated by some extra body model which carries the representational power, but are instead directly represented.

Keywords: embodiment, handedness, Ozma barrier, indexicals, bodily self-models

What is embodied cognitive content?

Embodied cognition is one of the major research movements of contemporary cognitive science. The central idea is that embodiment is a crucial feature for any understanding of the nature of cognition. According to many embodiment proponents representational and computational cognition arises or is driven from bodily, dynamic interactions of the cognitive system with the world and depends on the particular kinds of experiences that the system has, as a physically embodied system, with its environmental embedding, with which the system's internal activities are inextricably intertwined. In this respect, contemporary embodied cognitive science is closely linked to other fashionable paradigms such as *situatedness*, the view that cognition depends on the particular

* Philosophy Department, University of Bonn, email: lyre@uni-bonn.de

situation in which the cognitive system happens to be, *embeddedness*, the view that cognition depends on the particular environmental embedding of the cognitive system, and *dynamicism*, the view that cognitive systems should be viewed and studied by means of dynamical systems theory. By emphasizing and focusing on the bodily nature of our cognitive self, the modern doctrine of embodiment also shows remarkable similarity with certain strands of traditional phenomenology (most prominently Merleau-Ponty 1945).

Most obviously, embodied cognition raises various issues of eminent philosophical interest. One of the most sweeping debates concerns the question whether embodiment gives rise to non-representationalism about mental content. Such a radical claim has for instance been put forward by van Gelder (1995). Indeed one of the reasons to dismiss representationalism is that, at least for some tasks, it proves to be very useful to reconstrue cognitive systems as dynamical systems with a particular, environmentally driven state space dynamics, where no recourse to internal representations is necessary to explain the system's behaviour (it is also well-known that recurrent neuronal nets und dynamic systems with real-valued system quantities can essentially be mapped onto each other). However, the majority of authors still holds more moderate views (compare, for example, Bechtel 1998 or Clark 1997). Accordingly, the attitude of this paper will also be a representationalist one, but prospects of the particular nature of embodied representation will enter center stage in our discussion in the final part of the paper.

Another remarkable motive of embodiment is the thesis that cognitive systems are not confined to their traditional boundaries of skin and skull, but rather extend into those parts of the environment which are relevant to trigger the dynamic activity of the cognitive system. Under the supposition that mental content supervenes on the entire cognitive system, the mind may also be considered as extended, which leads to a new variant of content externalism, dubbed "active externalism" according to Clark and Chalmers (1998). And here an even more radical thesis is lurking in the back: the thesis that, eventually, the very idea of a cognitive core system has to be given up. But although this is certainly a highly interesting line to follow, we will not delve into these issues here but rather adopt the traditional view that talk about cognitive systems or agents is justified – and that those systems are, at least for all practical purposes, best viewed as being confined by their bodily boundaries.

The motivation for this paper is not only to follow the externalist view about contents that supervene on the embodied cognitive system (traditionally: the cognitive core system plus body), but to ask the more intriguing question whether *genuine embodied*

cognitive contents possibly exist. That is to say, are there cognitive contents, which solely exist in or can be represented only by truly embodied systems? Can any examples of such contents be given? This is the kind of questions which we like to follow here. And as we will see in a moment, the phenomenon of handedness – especially the which-is-which content of handedness – provides an intriguing example worth discussing.

Handedness and the Ozma barrier

In order to get a proper intuition about the which-is-which content of handedness we should start with a problem from technical communication theory, for which the recreational mathematician and author Martin Gardner has coined the name "Ozma problem": Consider the setting of the SETI project, the search for extraterrestrial intelligence by sending radio signals in outer space. Suppose further SETI would be crowned with success (what an incredible supposition!), then the following subtle technical problem arises (amongst many others, certainly far more urgent problems).[1] In Gardner's own words:

> *Assume we have already established fluent communication with Planet X by means of a language ... and by the use of pictures. We have asked them to scan their rectangles from "top to bottom" and from "left to right." There is no possibility of their misinterpreting what we mean by "top to bottom." "Top" is the direction away from the center of a planet, "bottom" is toward the planet's center. "Front and back" is no problem either. But having established the meanings of up, down, front, back, how do we make clear our understanding of that third pair of directions, left and right? How can we be sure, when we transmit a picture of, say, what we call a right-handed helix, they receive a picture of a helix with the same handedness? If they have taken "left to right" in the same sense that we use the phrase, the pictures will match, but if they are scanning the other way, our picture of a right helix will be reproduced on Planet X as a left helix. In brief, how can we communicate to Planet X our meaning of left and right?* (Gardner 1964, p. 166–167)

This is indeed a puzzling question – and it turns out that it has no positive answer! But perhaps, at first sight, it is not obvious why there is no problem with top-down and front-back, while there is one with left-right. The reason is that the first two are just conventional. Suppose we want to tell our friends from Planet X how to produce dice. After telling them how to produce cubes (objects with a certain number of edges and vertices and with a certain discrete rotational symmetry), we must number the faces

[1] A forerunner of SETI was the Project Ozma, named by it's director Frank Drake in admiration of Frank Baum's Oz stories and one of its fancy characters, the Long Eared Hearer. Hence Gardner's name-giving.

from 1 to 6. We tell them that opposite faces add up to seven. This implies that at one vertex the faces 1, 2 and 3 intersect. So we ask them to choose one face at random and call it 1 (and the opposing face 6). We then ask them to number one of the four neighbouring faces of 1 as 2. Hence, top-down and front-back is no problem. But how are we to tell them how to choose the faces for 3 and 4? There are two possibilities left and they mark the difference between clockwise and counter-clockwise ordering and, hence, between a right-handed and left-handed die. This alternative marks a real physical difference and isn't a mere conventional matter any longer. The Ozma problem now consists in the fact that there is no way of telling our friends the difference between right and left. Of course, they might establish a right-left convention by their own. But this doesn't doesn't alter anything about the real physical difference between the two dice orientations. This difference means that if, by some unexpected technological progress, we and our friends are able to meet in some later future, there will be only a fifty-fifty chance that their dice and ours are entirely alike. Now consider pictures: We may first align our convention about top and down in the way Gardner proposes (since pictures are two-dimensional, the question of front-back doesn't bother us her). But note that it is really just a convention, for even if they produced their pictures upside down and we would met some day, a simple rotation would make their pictures and ours alike – in case of all the symmetrical pictures! For asymmetrical pictures, however, the fifty-fifty chance of difference in mirror symmetry remains.[2] And there is indeed no way of communicating this difference by means of a usual data channel such as a pulsating radio signal.

This negative result holds true at least under two provisos. Of course, the crucial point about the Ozma problem is that any serial data channel, in principle representable as a string of 0's and 1's, is *itself* not a handed object. But it must also be excluded that there exist some other asymmetric, handed objects which we and our friends from Planet X can observe in common. Otherwise it would be just easy to match our left-right-conventions by means of an ostensive definition ("The spiral galaxy in sector XYZ of the universe is what we call left-handed"). By way of contrast, giving ostensive definitions is exactly the way – and apparently the only way – we define our left-right-conventions.

The second proviso is controversial and has to do with modern physics. Since 1957 it is well-known that parity P, a spatial transformation which in 3D is equivalent to mirror reflection, is not conserved. We may therefore perform P-violating experiments, which

[2] In more technical terms, the difference we observe here is the one between the continuous rotational symmetry group in space (with transformations connecting top-down and front-back) and the discrete mirror symmetry (connecting left-right).

can be used to define left- and right-handedness on the basis of law-like fundamental processes. And of course, given that our friends on Planet X have already discovered the laws of weak interaction, we may just ask them to perform such an experiment in order to match our conventions.

What makes this proviso controversial is a two-fold objection: First, in order to transmit the left-right-convention it is still necessary for our friends on Planet X to refer to some handed object (the outcome of a P-violating experiment). Of course in this case they won't use the same token, but at least the same type. This is what Martin Curd (1984) has dubbed "ostension at a distance". Second, the story about P-violation is actually far more complicated: In order to use a P-violating experiment as some kind of operational definition of a particular handedness, it is necessary that a convention about the charges of the elementary particles involved is established. Otherwise, and given the physical invariance of the combined operation CP (where C means charge reversal), the problem is just shifted (perhaps our friends consider "+" as "–"). Curiously, however, even CP is violated. Since 1964 physicists consider only the combined transformation CPT (where T means time reversal) to be conserved (deep mathematical reasons from quantum field theory speak in favour of this assumption, which wasn't refuted by any experiment so far). Hence, in order to establish P- and C-conventions with our friends on Planet X in terms of CP-violating experiments, we must presuppose a common T-convention before. At this point one might argue that the very possibility to communicate (in any rudimentary form whatsoever) already presupposes the same direction of time between the communicating partners and that, in this sense, the Ozma problem can finally be overcome. But, still, this is far from being a clear-cut result.

In any case, for the purposes of our following considerations, neither the first nor the second proviso can be applied, since we will discuss the "information transfer" between the outer world and internal models of cognitive systems only – and not between communities of intelligent beings). In this case, however, the Ozma problem has no solution. Since, in what follows, we will make extensive use of this insight, we may rather speak of it as the "Ozma barrier". It is precisely the information barrier which stems from the unsolvability of the Ozma problem under the two mentioned provisos.

Handedness: as such and which-is-which

As is well-known, handedness was for the first time introduced as a philosophical topic by Kant in his essay "On the First Ground of the Difference of the Regions in Space" (Kant 1768). For Kant the existence of left- and right-handed objects (which instantiate pairs of, as he called them, incongruent counterparts – objects which are alike in any

properties of their parts but still different in the sense that they cannot be enclosed by the same surface) proved that space does not depend on relations between things, but rather on relations between perceiving subjects and things. In his pre-critical 1768 essay Kant took handedness as showing that space is absolute and not relational, in his later critical writings he took it as showing that space is ideal or transcendental.

John Earman (1991) has claimed that P-violation provides a modern and refined variant of the Kantian challenge to spacetime relationalism (see my 2005 for arguments against this claim). While all this leads to various philosophical questions in their own right (mainly from epistemology and philosophy of physics; cf. Cleve and Frederick (1991) for a fine collection of articles about the philosophy of handedness), we have not yet encountered the connection to our overall question about genuine embodied contents. We should, however, notice the fact that there are two issues about handedness between which one must carefully distinguish:

1.) The nature of handedness as such,
2.) The nature of a particular handedness or which-is-which handedness.

In Kant, for instance, this distinction is notoriously unclear. It is a matter of controversial exegesis to find out whether Kant bound his arguments on either the nature of handedness in the first or in the second sense. For our purposes the second case is of particular interest, for it is only this case which is affected by the Ozma barrier. The reason is that it is perfectly possible to transfer the meaning of handedness as such over some data channel. We might for instance simply communicate the mathematical definition of a spatial reflection as an operation with two possible outcomes. It is as easy to communicate such a definition as any other mathematical definition whatsoever (and we do not doubt, for the purpose of this paper, that such transfers are possible in general).

It is therefore one thing to be a cognitive system which can be attributed the meaning of handedness as such (case 1) or to attribute a which-is-which content of handedness (case 2), henceforth abbreviated WWCH. We, ourselves, obviously have both meanings at our cognitive disposal: I do know what handedness as such means, but I do furthermore also know, what right or left in itself means.

Embodiment and spatiotemporal indexicals

What is the reason for our disposal about WWCH? Obviously, the reason for our possibility to distinguish left and right lies in the crucial fact that our physical body is mirror asymmetric. “Left is where the thumb is right” as the saying goes (or, more

seriously: left is where thumb and index finger shape an `L'). In this sense WWCH seems to provide a very good candidate for our overall question about genuine embodied contents. Indeed: suppose our friends on Planet X were living in a perfectly mirror symmetric part of the universe and they, themselves, albeit being embodied cognitive systems, have perfectly mirror symmetric bodies. Obviously in this case there is no chance to attribute a WWCH to them.

Note, however, that there is no principled reason why they shouldn't be able to understand the idea of handedness as such. They may have intellectual insight into the mathematical definition of oriented spaces and the like, but in the absence of handed objects being instantiated in their real surroundings they nevertheless lack the possibility to prefer left to right or vice versa (and they cannot use the L-trick since their letters have no intrinsic orientation). By way of contrast, our way of grasping WWCH apparently stems from our possibility to refer to our own body as an asymmetric, handed object. So WWCH is indeed essentially bound to some asymmetric embodiment – and it is certainly an impossibility in a disembodied cognitive system.

But let us push the issue a bit further. Surely embodiment comes in different varieties. In all day life we permanently rely on our bodily performance. Gestures, facial expressions or intentional actions in general are of that kind. They are not only based on our bodily performance but help us to transfer meaningful information. And isn't this quite generally true for any case of indexical reference? Isn't WWCH just one among many examples of indexical content?

Indexical content is generally understood as context-dependent referential content, where the context of an indexical expression (such as 'I', 'here', 'now') is construed as incorporating the speaker and the speaker's time and place. And quite naturally, as far as spatio-temporal indexicals are concerned, it is necessary to provide the speaking system with the possibility of locating it in space-time. In this sense, of course, any spatiotemporal indexical presupposes an elementary embodiment – some corporeal physical system.

Note that the story might be different for attitudes de se. Authors like Lewis (1979) refer to semantic content that is irreducibly de se in the sense that it relates in some essential way to one's own self and can therefore not be transformed into content de dicto or de re. While it seems to be a presupposition for a cognitive system to develop a "self" – however this can be done – it is not necessarily the case that such a self is part of an embodied system. This is different for spatiotemporal indexicals.

There exists, however, a longstanding debate whether the reference of indexical terms can generally be captured by some sort of descriptive content or not. Although meanwhile a certain consensus in the philosophy of language community is gained that this is not the case (compare for instance Perry 1997), it seems hardly possible to give a strong "proof" of this statement. This is precisely the reason why WWCH is of particular interest and which distinguishes WWCH from other spatiotemporal indexical contents, since, as should be clear from the our foregoing discussion, in case of WWCH such a proof is actually provided by the existence of the Ozma barrier! Indeed, WWCH is non-descriptive in nature and might as such be considered a *proven indexical* – perhaps even the only indexical of this strong sort.[3]

Embodiment, disembodiment and bodily self-models

We have seen that spatiotemporal indexical content is necessarily bound to embodiment and that the indexical nature of left and right can moreover be considered proven from the very existence of the Ozma barrier. Therefore the conclusion suggests itself that WWCH is a genuinely embodied cognitive content, a cognitive content which can be attributed – and for which we have a proof that it can be attributed – to an embodied cognitive agent only.

But isn't it nevertheless conceivable to have a WWCH without being embodied? What about the simulation of a body, a virtual body, as opposed to some real, materialized body in 3D space? In order to follow this line of thought we must find out whether and how it is possible at all to implement mirror-asymmetric objects within a virtual space. In principle, this doesn't seem to provide any serious problem. Consider, for instance, some arbitrary string of 0's and 1's as a representation of a 1-dimensional virtual space, then the sequences 1010010 and 0100101 provide simple examples of incongruent counterparts. In a 3D-world we may accordingly give the coordinate sets of the surfaces of a pair of hands. As a pure convention we may then, in a second step, call the first created object right-handed and its counterpart left-handed. And this simultaneously means that we have introduced a which-is-which meaning of left and right in the virtual

[3] The fact that P- or CP-violation may perhaps help to solve the Ozma problem for communication settings in our 3D world (as described above), does not affect our statement about the non-descriptive nature of WWCH, since a solution by means of P- or CP-violation is based on the pure contingent fact of the instantiation of the which-is-which handedness as a natural property of certain particles. But we should not confuse WWCH with that property. Whether nature realizes not only handedness as such, but also left- and right-handedness in the which-is-which sense, is a purely empirical question . Surely, as things stand, this is the case because of P-violation in modern physics, but even if this were not the case, we must nevertheless attribute a WWCH to ourselves simply because we possess an asymmetric body.

world. We may henceforth relate any further handed objects to this convention. So it seems possible to have handedness as such as well as which-is-which handedness as virtual properties. What is, however, impossible is to align our convention between left and right with the convention made in virtual space. This alignment is blocked by the Ozma barrier.

Imagine, now, a proponent of good old fashioned AI, who believes that we may reduce cognition to the processing of a universal Turing machine. He will of course claim that it is possible to have the body and all its supervening contents of proprioception reduced to a gigantic program. Let's call it PROGRAM. What PROGRAM thus emulates is nothing but a self-model of the body – not to be confused with the body itself. However, unless this conception doesn't collapse into solipsism, that is insofar as the GOFAI proponent also believes that the thinking PROGRAM has thoughts about the world outside, he soon or later faces the Ozma barrier as a problem of aligning inner and outer conventions about the WWCH.

But what exactly is the problem here? What PROGRAM does is of course not only to emulate its own body, it will simultaneously produce and possess a representation of the environment. Hence PROGRAM will integrate its corporeal self-model into a global world model – and it does of course pose no problem to align left-right-conventions between those two models, since they are part of one and the same virtual space, the program space of PROGRAM. Likewise, we may think of PROGRAM as a brain-in-a-vat. While there is no question about its internal consistency of its left-right-representations, let's call them L* and R*, it is not at all determined how they relate to left and right, L and R, as properties in the real world. We have, in fact, encountered a perplexing case of indeterminacy of reference.

Surely, the problem of reference is what vat-brain scenarios are all about (Putnam 1981). How do TREES* refer to TREES in the real world? Questions like these are at stake in debates about sceptical scenarios. But note that our problem is still a bit different (and perhaps deeper). Even in case of a solution to the problem of reference for vat-brains, there is no principled way to establish the reference from R* onto R as opposed to L or vice versa. The reason for this lies, of course, once again in the Ozma barrier.

Hence, our case about handedness brings a remarkable difference between the classical representationalist approach and embodiment to light. The doctrine of embodiment, at least in its stronger tenets, should leave no room for a body model over and beyond the

body itself.[4] There is no need to have an internal representation of what is already part of the cognitive system – the body itself. There is thus no need for surplus representations R* and L* of R and L. Left and right are *directly represented* by and within the body itself. This is, after all, the key idea of embodiment. The body is its own representation – and such a direct bodily representation is representation enough.[5]

A somewhat different but perhaps also intriguing way of putting the point is to compare it with the issue of qualia and spectrum-inversion scenarios. Some might rather think of our inner experience of left and right as qualia instead of mental content. Let's assume they are right (though I don't think so). By decoupling L* from L and R* from R we may of course have an inversion – a permutation of the internal R* and L* – without any relation to L and R in the real world and, accordingly, without any relation to the left hand side and right hand side of one's body. Our argument in favour of a directly embodied representation now transforms into an argument for left and right as directly embodied qualia by raising the usual Wittgensteinian objections against the presupposition of entities, in this case R* and L*, that can be changed without changing anything observable. Again, no such qualms occur in case of a direct embodiment.

The case of WWCH thus helps to clarify and highlight one of the central features of embodied cognition as opposed to more classical approaches. It can be seen as a case of a genuine embodied content in the sense that within a classical representationalist framework a peculiar problem of indeterminacy of reference occurs, since due to the Ozma barrier there is no way to fix the reference from R* onto R and L* onto L or conversely. No such indeterminacy, however, arises in the case of a direct embodied representation of left and right.

To be sure, this is not at all any sufficient evidence against the possibility of an internal body model or against specific forms of representationalism. Our case does, however, highlight the costs and consequences of such a view. Beings with an internal body model stand in danger of being referentially decoupled from the world as far as genuinely embodied cognitive contents such as WWCH are concerned. Such beings are referentially "trapped", as it were, in their internal bodily self-model. By way of contrast, embodied cognition offers an alternative to such a view, which would otherwise lead to a strong and principled variant of meaning anti-realism.

[4] Cf. Gallagher (2005) and Metzinger (2003) for different views and versions of self-models and body-models.
[5] As indicated in the beginning, for some this is even a reason to give up representationalism altogether. For our consideration, however, there is no need for such a strong philosophical claim.

Concluding remarks

Let's sum up. Our considerations have, firstly, shown that WWCH provides a case of what might be called a "proven indexical" due to the existence of the Ozma barrier, the impossibility to communicate the which-is-which meaning of left and right. From this it follows that WWCH is genuinely embodied in the sense that it cannot exist in a disembodied or even only mirror symmetrically embodied cognitive system. We have then discussed the question whether a simulated body wouldn't be sufficient. And although the answer to this is in the affirmative, we discovered another difficulty. While it seems natural for classical representationalist accounts to assume an internal representation of the body, a bodily self-model, as an instance onto which motor control and regulation act, the introduction of such a self-model creates a perplexing problem of referential underdetermination in case of genuinely embodied cognitive contents such as WWCH.

One final remark: While perhaps indexicals in general provide cases for embodied contents, it seems hard to find examples of genuinely embodied contents in the above construed strong and "proven" sense. Is WWCH perhaps even the only one? Compared to the which-is-which handedness as a natural property, the which-is-which charge and the which-is-which direction of time are on the same level (and, in fact, any which-is-which property connected with the three discrete fundamental transformations in physics: P, C, T, and their combinations). Is there, analogously, a which-is-which meaning of charge and a which-is-which meaning of the direction of time? As far as the case of charge is concerned, there is certainly nothing like that in human beings, since we simply do not have experiences of charges, not even charge *per se*. But it might very well be a positive example in charge sensitive beings as, for instance, electric fishes. The case of the which-is-which meaning of the time direction is more complex. It might be an interesting line to follow whether our knowledge about the which-is-which meaning of the time direction, insofar such a knowledge truly exists in us, is essentially bound to the irreversible life development of our body. This consideration is, however, complex and controversial for at least two reasons: it is, first, not as obvious as in the case of WWCH, whether we really have a which-is-which meaning of the time direction at our disposal, and second, complicated questions about the nature of irreversibility also come in – questions, which provide stuff for more than another paper. We may thus conclude that, as far as a clear cut example of a genuine embodied cognitive content is concerned, WWCH is indeed distinguished.

References

Bechtel, W. (1998). Representations and Cognitive Explanations: Assessing the Dynamicist's Challenge in Cognitive Science. *Cognitive Science* 22(3): 295-318.

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

Clark, A. and Chalmers, D. (1998). The extended mind. *Analysis* 58(1): 7–19.

Curd, M. (1984). Showing and telling: Can the difference between right and left be explained in words? *Ratio* 26: 63–66. (In Cleve and Frederick 1991).

Earman, J. (1991). On the other hand...: A reconsideration of Kant, incongruent counterparts, and absolute space. In van Cleve and Frederick 1991.

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

Gardner, M. (1964). *The Ambidextrous Universe*. Basic Books, New York. (Third revised edition: *The New Ambidextrous Universe*, Freeman, 1990).

Kant, I. (1768). Von dem ersten Grunde des Unterschiedes der Gegenden im Raume.

Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review* 88: 513–543.

Lyre, H. (2005). Metaphysik im „Handumdrehen" – Kant und Earman, Parität und moderne Raumauffassung. *Philosophia Naturalis* 42(1): 49-76.

Merleau-Ponty, M. (1945). *Phénoménologie de la Perception*. Gallimard, Paris.

Metzinger, Th. (2003): *Being No One. The Self-Model Theory of Subjectivity*. MIT Press, Cambridge, MA.

Perry, J. (1997). Indexicals and demonstratives. In Hale, B. and Wright, C., editors, *A Companion to the Philosophy of Language*. Blackwell, Oxford.

Putnam, H. (1981). *Reason, Truth and History*. Cambridge University Press, Cambridge.

van Cleve, J. and Frederick, R. E., editors (1991). *The Philosophy of Right and Left*. Kluwer, Dordrecht.

van Gelder, T. (1995). What might Cognition be if not Computation? *Journal of Philosophy* 91: 345-381.