Does the Brain Represent the World? Evidence Against the Mapping Assumption

Astrid von Stein
Institute of Neuroinformatics, University/ETH Zurich
Winterthurerstr. 190, CH-8057 Zurich, Switzerland
Email: astrid.vonstein@ini.phys.ethz.ch

Abstract · Whatever approach regarding internal representations, the idea was always that of a mapping of an outside world, more or less successfully performed by our cognitive apparatus. In the following we want to develop a principally different approach where representation is no more considered any kind of mapping of a predefined external reality, but simply as stabilities in the coupling between organisms and their local environment. Since this kind of representation evolves in the dynamic interaction with the environment it is a fundamentally active process of construction and not a passive mapping. Neuroscientific and psychological evidence favor this concept over old concepts on purely bottom-up mapping of the environment.

Concepts on Representation and Their Problems

By introspection we are convinced to have a picture of the outside world in our brain; therefore much neuroscientific work has been trying for decades to establish correlates of this mental representation. First big steps in this direction have been the discovery that the sensory world is mapped onto different parts of the cerebral cortex, and even further that within each modality, different parts of the perceived world are represented in a topographic manner (Fritsch & Hitzig 1960, Mountcastle 1957). Hubel & Wiesel's findings of cells in visual cortex that respond selectively to certain aspects of a visual stimulus gave evidence that a visual object might be reconstructed step by step from basic features such as orientation lines, angles etc. in a hierarchical process. At the end of the hierarchy there would be a cell that selectively represents the whole object. Thus these findings gave rise to the notion that cells in the cerebral cortex would represent environmental entities.

Although appealing, this concept today is hard to sustain already from a neuroscientific perspective (for arguments againts single cell coding see e.g. Braitenberg 1991, Dudai 1989, etc.). Additionally, many psychophysical results are difficult to explain on that basis. Therefore in recent years a new concept of distributed representation has been developed that better accounts for both, the neuroscientific and the psychological data (e.g. Rumelhard 1986). Today both theoretically and empirically most of the models on cortical representation assume that there is at least a number of cells (a population) that represent an environmental entity. The approaches differ in their conception about the functional formation of these populations: whether the population is defined by pure activity in response to a stimulus (population coding) or whether the spatio-temporal structure of activity within a group of neurons plays a role (Abeles 1991). Synchronous activity among distrubuted neurons has been proposed to bind them into a functional cell assembly (temporal coding; König et al. 1995, Singer 1993, von der Marlsburg et al. 1986). However evidences seem to converge on the basic conclusion that entities are not represented on a single locus but in distributed functional assemblies . This approach also seems more adequate to convey the biological function of representation. Representations are not predefined but they have to be learned. Therefore the environment may change and still the cerebral representation may be adapted; new representations may be learned. Further environmental entities are not as fixed as in single cell representation; what defines an entity are the Gestalt laws such as common motion, common disparity etc. And indeed, the cortical network seems equipped to encode and decode environmental stimuli according to such Gestalt laws (Singer 1993). Besides several problems that have been solved,

Understanding Representation in the Cognitive Sciences
Edited by A. Riegler, M. Peschl, and A. von Stein. Kluwer Academic/Plenum Publishers, New York, 1999

there are several challenges even to this new concept of representation. Increasing evidence shows that the cerbral network is not just a feedforward network where the outside-space can be easily mapped onto, but that the cerebral architecure is extremely recurrent on all levels of processing. Both the local network within one cortical area and the interareal network between different levels of the hierarchy have multiple reciprocal connections (Douglas & Martin 1998). Thus processing is not likely to consist of one flow of information from the outside world to the internal representation in cortex. It rather seems that the internal dynamics of the network must have an equal impact on cortical processing as the input from outside. Indeed, physiological experiments show that the receptive field properties of neurons in V1 are altered by the activity of feedback connections from higher cortical areas (Bullier 1996). Although there are ways to represent patterns even in recurrent neural networks (e.g. Amit 1989, Elman 1990), on a conceptual level these findings still provide a deep challenge for our concept of representation as a mapping of an environmental stimulus onto some kind of internal representation. The problem can be posed in the following way. In distributed representation it is assumed that an entity is represented by the activation matrix of the network of neurons. The history of correlations embedded in the connectivity matrix (learning) guarantees that each input pattern chooses a certain pattern of activation in the network, which is therefore representing this input. A pattern of activation of n elements can be depicted as a point in an *n*-dimensional vector space. Each entity ist thus represented by a point in vector space. State transitions in this system can be easily studied if depicting them according to automaton theory in fig.1: The letters are the states of the system and the numbers the input to the system. The figure demonstrates a fundamental property of dynamical systems: the actual state of a system depends not only on the input, but also on the *previous state of the system*. This, however, leads to a severe problem in representation of environmental entities. How can an assignment between an environmental entity and its (representing) activation state in the network be guaranteed if the induced state depends not only on the input but also on the internal state of the system at the moment the input arrives (Peschl 1994)?

In a previous paper (von Stein 1994) we proposed a *reset mechanism* towards a reference state prior to input entering the system as a solution to the problem. If each input (perception) starts from this reference state, an unambiguous assignment is possible. Since in human EEG an episode of alpha rhythm prior to an expected input is often reported, we proposed the cortical alpha rhythm to function as such a reference state in the brain. Reticolo-thalamo-cortical loops involved in alpha generation and blockage due to attentional processes might be the operators for the postulated mechanism. However a reset mechanism is only a partial solution to the problem in cases where the input is expected (attentional processes). Another much more basic explanation is that our concept of representation understood as a reference between an environmental entity and some cerebral correlate, *is wrong*.

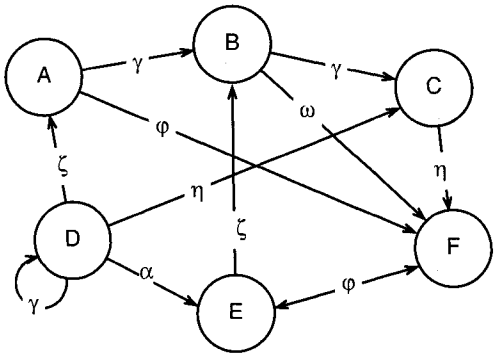

Figure 1

A Different Approach to Representation

I am going to develop a solution to the problem by questioning the nature of representation as a mapping process of a predefined external reality. As a first step I am going to investigate the nature of the biological function of representation. Defining the function of representation may help to elucidate the workings of the organ supposed to subserve that function, the cerebral nerve system (CNS).

The basic assumptions are lent from constructivism (Maturana & Varela 1980, von Glasersfeld 1985, etc). According to this theory, living beeings are entities that are capable of existing in a constantly changing environment without loosing their internal organization and structure. To do so, they have to provide several mechanisms to either exchange substances with the environment or react on the environment in changing it. Thus organisms

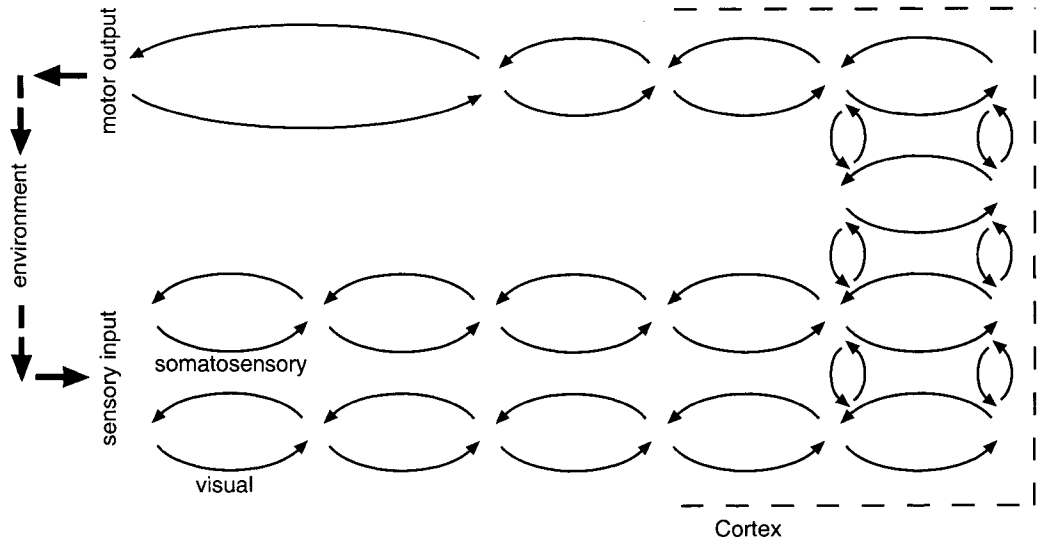

Figure 2

are coupled with their environment in a dynamic interaction maintaining a stable equilibrium. Only if the coupling between environment and organism will be successful in this respect, the organism will survive. Systems that are capable of this stabilizing process are called autopoietic systems (Maturana & Varela 1980). All living beeings are autopoietic systems. The mechanisms to provide stable coupling with the environment however are different. Simple systems are mere input-output systems, where the environmental pattern that activates a sensor and the given motor response are hardwired into the system; the usefullness of a coupling is then established by evolution. Complex systems, on the other hand, have developed a mediator between input and output, the central nerve system (CNS). This highly connected structure situated between the input organs (sensors) and the output organs (effectors) seems to enlargen the space of possible sensomotoric couplings and make it more flexible to different environments. Not one single substance automatically causes an avoidance or attractance response, but the patterns that elicit responses are more complex, consisting of a combination of input activations, with what is forming an actual relevant "pattern" is learned and not hardwired. Additionally, the *reaction* to this pattern is not prewired but learned, and flexibly dependent on the simultanous information arising from other input channels. This gives the system far more degrees of freedom both on the sensory side and on the output side. Increasing the complexity of interactions seems to be one of the evolutionary goals of a CNS. An autopoietic system with a CNS seems better equipped to survive in constantly changing environments and to survive individually instead of evolutionary. This CNS will help it to constantly change strategies in the coupling with its environment to provide stability in this coupling. Considering the CNS from this perspective helps elucidating the problem of representation: the purpose of the brain does not seem to *represent* the world but rather to provide means of a stable coupling with the environment. From this perspective the afore-mentioned pecularity of its recurrent architecture does not at all create a problem but contrarily seems perfectly suited to subserve its function:

The CNS as the mediator between environment and internal millieu is a densely packed medium of millions of little sensory–motor loops. Not only is the whole system from the sensory receptors to the muscles a big sensory motor loop with the environment, but at each step of processing there is a connection backwards towards the periphery thus creating a small loop instead of a pure feedforward connectivity (see figure 2). Thus the whole system can be viewed as a complex net of interwoven sensory–motor cycles (micro loops). The goal of the system is to keep stable couplings with the outside despite a permanently changing environment. To do

so it has stored old successful couplings in its connectivity matrix and reacts with adequate output towards destabilizing inputs from the environment according to the internal dynamics of the thus created network. Any successful reaction is again stored in the sensory–motor matrix, thus creating stabilities in the network that reflect organism-related stabilities with the environment. These stabilities, emergent phenomena of network dynamics reflecting the history of couplings, are the basis for what we perceive as "representation" of the environment. Thus, interpreting representation in this way solves the reference problem that arises when dealing with a recurrent architecture like the cortex. There is no need for a fixed reference between environment and internal state of the network, but only for a stable coupling between both, where the position of both might actually change. This representation is no form of mapping of a given environment, nor is it only a construct of our mind (as radical constructivism says): it is one part of a circulatory process where constant parts in this process, both on the sensory side as on the motor side, are represented. Therefore, if one tries to define representation by any kind of reference between an external world and an internal correlate—whether this correlate is representing neurons (grandmother neurons) or states of the cortical network—this approach principally must fail: it tries to correlate entities in the environment with cortical correlates, whereas entities are not outside, nor inside, but only defined by the interaction between both.

Evidence from Neuroscience and Developmental Psychology

Several findings from neurophysiology and anatomy make more sense in the light of this interpretation. The extreme reciprocity of connections giving the system a strong internal dynamics is not a problem but an actual purpose. Processing of a stimulus is not a passive propagation from the periphery but an active process of holding or creating a stable equilibrium: therefore this process might be initiated either by a change in the environment (i.e. what we call "a stimulus") or by a change within the system, i.e. the internal dynamics. If the environment changes, the organism will react onto it. If the internal millieu changes, the organism will act onto the environment. Both pertubations to the system lead to a reaction, thus pushing the system towards a new stable state. Therefore, both internal state and external input have an equal impact on cortical processing. The fact that processing is fundamentally *active* is demonstrated by the fact that the cortex is not *quiet* in the absense of sensory input. Rather is there a constant background activity, or a resting activity in the cortical network at any time. This may be evidence that sensory–motor loops are constantly active, checking and updating whether the environment has changed or whether it does present any interesting input that might fit to an actual internal instability. This permanent activity within the sensorimotor loops might reflect the dynamic equilibrium with the environment as described before. If this is the case, background activity however should not not just be noise within the system but reflect specific processes of interaction. Indeed it has been shown that classical background activity in the EEG (the so called alpha rhythm or other low-frequency rhythms) is not merely noise but reflects specific mental processing: highly specific patterns of cortical dynamics within the frequencies of the classical resting-rhythm have been found in the absense of visual stimuli such as during mental imagery or working memory in humans (Petsche et al.1996, Sarnthein et al. 1998). Also, similar specific patterns of slow-frequency interactions have been found in intracortical recordings in cats (von Stein et al. 1996). Interestingly, rhythmic activity driving the resting rhythm (alpha) have been found in the cells of layer V that project to the superior colliculus (occulo–motor system) (Silva et al 1991); this fits to the idea that the slow frequency rhythms of background activity may be the correlate of activity within the complex network of micro sensory–motor loops. Further, it has been shown that the background activity that is present before a stimulus enters the system, radically influences processing of that stimulus (Arieli et al 1996). Influences of the activity of the cortical network on processing of an external stimulus, specifically via the top down connections from higher hierarchical areas, has also been shown with current source analysis in monkeys (Cauller 1991).

Additionally, studies on the development of object representation during childhood show how these sensory–motor loops might have formed, and how they later interact to form complex stabilities—"percepts" or "concepts". Piaget (1959, 1970) has shown that in the very early stages of child develop-

ment each sensory channel seems to work as an independent device. For example in the visual system, if light is entering the system, the childs eyes will move towards it. Thus an input to the retina activates the occulomotor system and induces a response, very much like in a simple stimulus response system. Similarily for the other modalities, if an object will touch the childs fingers it will grasp it, if something touches the childs mouth it will suck it. However that these sensory–motor loops are not simple stimulus–response devices becomes evident if we observe what happens when there is *no* source of light to the retina. In this case we would expect that an input–output device not to be active; this however is not the case. Instead, in moments without input, the childs occulomotor muscles constantly swift around until they catch some bright object. Piaget calls that "assimilation": it seems as if there is a mechanism working to keep the loop in permanent activation: either it is activated by an external stimulus moving across the retina or it is activated by the muscles moving the retina relative to the external world. Both cases lead to activity within the loop, and noone could define easily who was first. This situation reminds of the stable equilibrium described in the section above. On each part of the sensory peripheries, there is a permanent dynamic interaction between environment and organism within these first simple sensory–motor loops. Each of it guarantees that changes on either side are immediately detected leading to a reaction and thus a new stable state. In this way, the sensorimotor loops seem to both be ready for changes and to detect invariances. Thus, in the first behavior of children we can detect the activity of the process as described above.

Interestingly, these different sensory–motor loops are yet working completely independently. Thus, if the same bright object touches the childs finger activating the grasping response and enters the childs eyes activating the occulomotor response, both are not yet integrated. Only later integration will develop and the child will notice that it is not dependent on pure chance wether the object will fall into its hand but that there is actually a relationship between the visual channel having seen an object and the somatosensory channel having sensed an object. This relationship will help it to finally purposely guide behavior to aquire objects. This integration however is the first step of forming a common representation of an object: the different channels are no longer processed independently but have interacted and established an invariance, the object. Piaget describes several steps of the childs development towards these higher order representations. He terms the first basic sensory–motor loops "primary circulary-reactions", and describes behavior that gives evidence of secondary circulary-reactions, tertiary circulary-reactions etc. Thus it seems as if the first peripheral sensory–motor loops start to interact with sensory–motor loops towards the central nervous system etc. and finally interact with each other. It is easily conceivable that these different loops are equivalent of the growing connections between hierarchical areas, with feedback fibers growing step by step and being strengthened as simultanous activation of neurons from other sensory areas occurs simultanously. In a purely Hebbian sense, two sensory–motor loops from different sensory systems should become coupled if repeatedly activated by one object. If during exposure with the environment repeated interactions with an object occur in several channels, therefore finally the first "representations" are formed. In conclusion, what we finally perceive as "objects" in our mind are again stabilities in the interaction with the environment. However, opposed to the first peripheral stabilities that were pure stabilities within one micro sensory–motor loop, these stabilities include the interaction between both, the sensory–motor loops with their environment and the senory–motor loops within the cortex. Extracting invariances on this more complex level seems to be the final goal of this process. Several of such invariances will develop and help the organism to find stabilities in each given behavioral situation (adequate behavior).

Conclusions

On the basis of various arguments I tried to show that representation is not a mapping of predefined environmental entities onto cortical activation-states. From a theoretical perspective, in recurrent architectures like the cerebral cortex the internal dynamics of the network provides a bad medium for mapping because the actual activation-state is dependent on the previous activations state. The best solution to this problem of reference between external entity and internal state is to give up the concept of mapping. It seems more appropriate to define representation as a stable interaction between environment and internal state, with many different

solutions (substabilities within the network); the actual establishment of substabilities might give rise to the subjective experience of entities. Representation in this way is not a passive bottom-up mapping but an active interactive process between external requirements and internal requirements trying to stabilize on the most feasible solution for the organism. The actual neurophysiological data supports the notion of an active process, and of an interaction between bottom-up and top-down processing. Anatomical data shows that the CNS may be consisdered as a system of interwoven sensory–motor loops (micro-loops). Observations on child development show how such micro-loops have the tendency to create stabile equilibriums with the environment; they also demonstrate how they finally might interact during learning to form higher order stabilities (invariances) and thus create what we experience as representations of environmental objects.

References

Abeles, M. (1991) *Corticonics*. Cambridge: Cambridge University Press.

Amit, D. J. (1989) *Modeling Brain Function. The World of Attractor Neural Networks*. Cambridge: Cambridge University Press.

Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. (1996) Dynamics of ongoing activity/ explanation of the large variability in evoked cortical responses. *Science* 2(73): 1868–1871.

Braitenberg, V. & Schüz, A. (1991) *Anatomy of the Cortex. Statistics and Geometry*. Berlin: Springer.

Bullier, J., Hupé, J. -M., James, A. C. & Girard, P. (1996) Functional interactions between areas V1 and V2 in the monkey. *J. Physiol.* 90: 217–220.

Cauller, L. J. & Kulics, A. T. (1991) The neural basis of the behaviorally relevant N1 component of the somatosensory-evoked potential in SI cortex of awake monkeys: evidence that backward cortical projektions signal conscious touch sensation. *Experimental Brain Research* 84 (1991): 607–619

Douglas, R. J. & Martin, K. A. (1998) Neocortex. In: Shephard, G. M. (ed.) *The Synoptic Organization of the Brain*. Oxford University Press.

Dudai, Y. (1989) *The Neurobiology of Memory*. Oxford University Press.

Elman, J. L. (1990) Finding structure in time. *Cognitive Science* 14: 179–211.

Fritsch, G. & Hitzig, E. (1960) Über die elektrische Erregbarkeit des Grosshirns. *Arch. Anat. Wiss. Med.*, pp. 300–322. (Engl.: G. von Bonin (trans) In: *Some papers on the Cerebral Cortex*. Springfield, IL: Thomas, pp. 73–96.)

König, P. & Engel, A. K. (1995) Correlated firing in sensory–motor systems. *Current Opinion in Neurobiology* 5: 511–519.

Maturana, H. R. & Varela, F. J. (1980) Autopoiesis and Cognition. *The Realization of the Living*. Dordrecht, Boston, London: D. Reidel Publishing Company.

Mountcastle, V. B. (1957) Modality and topographic properties of single neurons of cat somatic sensory cortex. *Journal of Neurophysiology* 20: 408–434.

Peschl, M. (1994) Autonomy vs. environmental dependency in neural knowledge representation. In: Brooks, R. & Maes, P. (eds.) *Artificial Life IV.*. Cambridge: MIT Press.

Petsche H, von Stein, A. & Filz, O. (1996) EEG aspects of mentally playing an instrument. *Cogn Brain Res* 1: 115–123.

Petsche, H., Kaplan, S., von Stein, A. & Filz, O. (1996) The possible meaning of the upper and lower alpha frequency for cognitive and creative tasks: a probability mapping study. In: Basar, E., Lopes da Silva, F. & Hari, R. (eds.) *Alpha Processes of the Brain*. Boston: Birkhäuser.

Piaget, J. (1970) *Carmichael's Manual of Child Psychology*. New York: J. Wiley and Sons, Inc.

Piaget, J. (1959) *La naissance de líntelligence chez l'enfant*. Neuchatel, Switzerland: Delachaux et Niestlé S.A.

Rumelhart, D. E., McClelland, J. L. and the PDP Research Group (1986) *Parallel Distributed Processing*. Cambridge: MIT Press.

Sarnthein, J., Rappelsberger, P., Shaw, G., & von Stein, A. (1998) Synchronization between prefrontal and posterior association cortex during human working memory. *Proceedins of the National Academy of Science, USA, Vol. 95*, pp. 7092–7096.

Silva, R., Amitai, Y. & Connors, B. W. (1991) Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. *Science* 251: 432–435

Singer, W. (1993) Synchronization of cortical activity and its putative role in information processing and learning. *Ann. Rev. Physiol.* 55 (1993): 349–374.

von der Marlsburg, C. & Schneider, W. (1986) A

neural cocktail-party processor. *Biol.Cybern.* 54 (1986) 29–40.

von Glasersfeld, E. (1985) Einführung in den radikalen Konstruktivismus. In: Watzlawick, P. (ed.) *Die erfundene Wirklichkeit.* München: Piper.

von Stein, A., Chiang, C. & König, P. (1996) Expectancy driven synchronization between primary visual cortex and parietal cortex in cats. *Society for Neuroscience Abstracts.*

von Stein, A. & Peschl, M. (1994) Synchronization–Desynchronization. In: Eiselt, M., Zwiener, U. & Witte, H. (eds.) *Quantitative and Topological EEG and MEG Analysis.* Jena: Universitätsverlag Druckhaus-Mayer GmbH.