semantics consequent to the built–in evaluation criteria. A similar limitation is pointed out by Pfeifer and Scheier, who describe a "trade–o" between specificity and generality of value systems" ([8], p. 473): A very specific value system will not lead to a high degree of flexibility in behaviour, while a very general value system will not constrain the behavioural possibilities of the agent su"ciently.

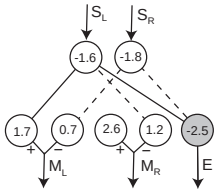
The common denominator of these di"erent issues raised by di"erent researchers is summarised in Rutkowska's question of whether a value system constitutes a "vestigial ghost in the machine" ([9], p. 292). A value system that applies pre–specified evaluation criteria to pre–specified sensory states to steer ontogenesis in a top–down manner, even if it guides the adaptation of real–time situated and embodied behaviour, is in itself a disembodied control structure. As such, it su"ers from all the problems associated with traditional disembodied artificial intelligence architectures, which have been pointed out many times (e.g. [2, 7, 8]): They are rigid and non–adaptive, their functionality relies on the intact functionality of dedicated input and output channels and they can only deal with scenarios that could be foreseen when they were designed.

2.3 The Only Good Ghost Is a Dead Ghost

The astonishing fact about value system architectures is that, despite the outlined disembodied nature of the value system, these architectures are very popular with researchers that share our concerns about situatedness and embodiment in the study of intelligent behaviour, and who are deeply sceptical towards classical symbolic approaches. For instance, Sporns and Edelman point out how TNGS models, through their increased flexibility, can overcome di"culties such as anatomical variations, which are "challenging to traditional computational approaches" ([10] p. 960). It is probably unquestioned that "Understanding Intelligence" by Pfeifer and Scheier [8], the very volume that advertises value guided learning, is one of the most important books to promote the situated and embodied approach.

changes in the cortex[5]. The bigger question to be asked in this context is: What can we deduce from such a correspondence ?

The fitness ( ))ωi


S_L
S_R
-1.6
-1.8
1.7
0.7
2.6
1.2
-2.5
+
−
+
−
M_L
M_R
E

We now return to the agent's value system. The estimator neuron M outputs $\rightarrow \approx 0$ if $_L = 0$. The reason for this is that during the entire approach behaviour $_L = 1$, and therefore $_L = 0$ implies that the light has not yet been located, which only happens in the beginning of the trials if the agent is far away from the light source. During the nearly straight path segments, $_L = = 1$, which leads to $\rightarrow \approx 0\ 5$, i.e. an intermediate estimate for an intermediate approach stage. While the agent cycles around the light source, $= 0$ and $_L = 1$, and the value system produces its maximum estimate, expressing that the light source has been reached. Notice also that the straight path segments which correspond to $\rightarrow \approx 0\ 5$ become shorter as the agent comes closer to the light. Therefore, even though the value system has just three modes of output, its evolution over time can express a more gradual change in distance, if averaged over a time window: The average output increases with decreasing distance to the light.

Another event worth discussing in the trial depicted in Fig. 2 (B) and (C) occurs after the last displacement of the light source (2800): As the displacement happens to bring the light source in the left visual field of the agent,

(A)

The presented results hopefully illustrate how these two options exclude each other: An "embodied value system" is a contradictio in adjecto. The existence of reciprocal causal links between value system and behaviour generating systems causes semantic drift of the value signal, which results in anarchy of development (see Sect. 4.2). But how could a value system not be embodied? Surely, we do not want to introduce magic meaning sensors or a magic master value system that ensures that the other value systems work smoothly. This smells too much of what Rutkowska calls "[b]uck passing to evolution" ([9], p. 292). If we struggle to explain the simple case without such scaffolding, the more abstract case will surely not become easier. The only way a value system architecture can work is a full embracement of the functional separation and pre–specification of meaning.

In the area of robotics, as shown in [12], we can design experiments rigidly enough to fixate meaning. But for an approach that aims at advancing past the stage of pre–specified motor programs, that refers to variable biomechanical properties in living organisms, the introduction of parts of the organism that are exempted from ontogeny, despite the constant material flux an organism undergoes, seems like a step backwards. It appears so inevitable that a random change would slightly change the context in which a value system is embedded, and the value–agnostic remainder of the organism would be unable to detect it or do anything about it. Furthermore, both in the area of biological modelling and in robotics, there is another unpleasant side–effect resulting from the introduction of disembodied and non–adaptive value systems: The impossibility of novel values. A rigid structure with a priori meaning can only work in situations that rely on phylogenetic constancies, the generation of new values in situations that our ancestors could not even have dreamt of asks for a different explanation.

We do not want to question that structures like the ones described as value systems exist in living organisms and that they play an important role in the adaptation of behaviour. In contrary, we think that the investigation of such mechanisms is important and intriguing. We plan follow–up experiments to the

entail, or even justify, the reduction of the respective fun