Can (Philosophical) Zombies Learn?

Nihat Ay[1,2,3], Carlotta Langer[1]

July 18, 2022

[1]Hamburg University of Technology,
Institute for Data Science Foundations, Hamburg, Germany

[2]Santa Fe Institute, Santa Fe, NM, USA

[3]Leipzig University, Leipzig, Germany

It is generally believed that truly intelligent agents should have the capacity of conscious experience. Integrated Information Theory (IIT) postulates that conscious experience is the result of information integration in the brain. We contrast this with the general conceptual foundation of the filed of embodied intelligence. There, various principles that underly intelligent behaviour have been revealed. One of these principles states that intelligent behaviour should exploit the potential of the agent's body and its environment for reducing the computational cost of the brain. This reduction is referred to as *morphological computation* and can be considered as the computation that the brain outsources to its surroundings. This notion of intelligence, however, creates a paradox in view of IIT. If an agent is so intelligent that it outsources most of the computation to the outside world then there will be no information integration within the brain. Thus, according to IIT, extremely intelligent creatures will not be capable of any conscious experience. Evolution towards intelligence will create zombies, a philosophical notion of an agent that generates complex and natural behaviours, as we do, but without being able to experience them consciously.

FIGURE 14 | Integrated information and morphological computation for the reactive control agents.
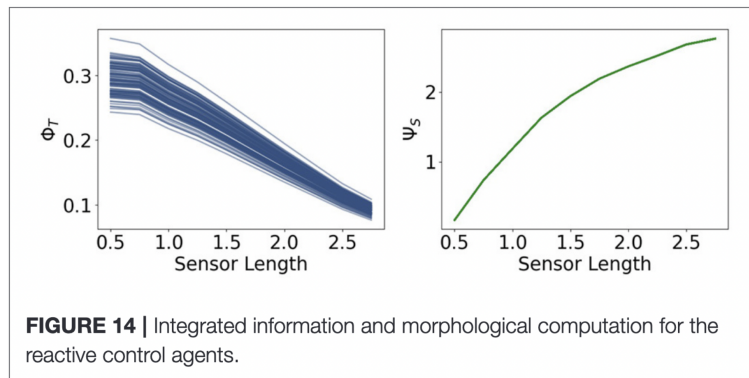
Figure 1: These plots demonstrate the antagonistic relationship between integrated information and morphological computation in an experimental setting. They are chosen as a sample from the article [LA20] which contains a much more detailed analysis of this relationship.

In this talk, we present a simple case study that supports the outlined intuition. More precisely, we use a measure of morphological computation [ZLA17] and a measure of integrated information [LA20] and show that optimal behaviour implies an antagonistic relationship between these

two measures [LA20]. The more morphological computation is involved the less information is integrated in the brain (see Figure 1).

Even though we do believe that intelligence, in the sense of outsourcing computation, ultimately creates zombies, we consider this as a paradox of our limited understanding of intelligence than a general law. As a solution, we propose that information integration is involved not only in the control but also in the planning of behaviour. That requires a world model as basis of deliberative action and learning. Zombies can well control behaviours based on given policies. The question is: Who gives them these policies? They will not be able to obtain such policies without a world model as basis of learning. This suggests, as an answer to our question in the title, that zombies cannot learn, which constitutes a working hypothesis of our ongoing research supported by a few initial results.

References

[LA21]   Carlotta Langer, Nihat Ay. *How Morphological Computation Shapes Integrated Information in Embodied Agents.* Frontiers in Psychology 2021, doi: 10.3389/fpsyg.2021.716433.

[LA20]   Carlotta Langer, Nihat Ay. *Complexity as Causal Information Integration.* Entropy 2020, doi:10.3390/e22101107.

[ZLA17]   Keyan Ghazi-Zahedi, Carlotta Langer, Nihat Ay. *Morphological Computation: Synergy of Body and Brain.* Entropy 2017, 19(9), 456; https://doi.org/10.3390/e19090456.