Simulation of Non-Primate Intelligence vs Human Intelligence vs Superhuman AGI vs Alien-like AGI

Schneider, Howard[1[0000-0001-8052-6448]]

[1] Sheppard Clinic North, Vaughan, Ontario, Canada
hschneidermd@alum.mit.edu

Abstract. The Causal Cognitive Architecture is a brain-inspired cognitive architecture whereby millions of neocortical minicolumns are modeled in the architecture as millions of navigation maps, capable of holding spatial features and small procedures. The Causal Cognitive Architecture 7 (CCA7), possessing the same properties of its predecessor of fully grounded, continuous lifetime learning, associative reasoning, full causal reasoning, analogical reasoning, and near-full compositional language comprehension, also possesses superhuman planning abilities and on a conceptual level can serve as a proxy for superhuman artificial general intelligence (AGI). A simulation of this architecture, and subsets of it, are used to model pre-mammalian and non-primate mammalian-level artificial intelligence (AI), human-level artificial intelligence (HLAI), superhuman AGI and links to a large language model (LLM) as a proxy for an alien-like (i.e., non-biologically-based) AGI. The models were tested on a compositionality problem with the best scores: superhuman >= HLAI > LLM > pre-mammalian/mammalian ($p<0.001$). Testing on a traveling salesperson problem: superhuman > LLM > HLAI > pre-mammalian/mammalian ($p<0.001$). These results indicate the need to consider intrinsic compositional and planning abilities in the development of AGI systems.
Keywords: Artificial General Intelligence (AGI), Superintelligence, Cognitive Architecture, Compositionality, Planning.

1 Introduction – The Evolution of a BICA from Associative Reasoning to Superhuman Intelligent Behavior

Brain-Inspired Cognitive Architectures (BICA's) are cognitive architectures inspired by the human brain [1]. The Causal Cognitive Architecture is a BICA which hypothesizes that the navigation circuits in the amniotic ancestors of mammals were duplicated multiple times resulting in the evolution of the neocortex [4, 5, 6]. Therefore, the thousands or millions of cortical minicolumns in the mammalian brain are considered to be thousands or millions of spatial navigation maps. The architecture does not tightly replicate the mammalian brain at the level of spiking neurons nor at the behavioral level but considers in a functionalist sense [7] what properties emerge from thousands or millions of such navigation maps in a cognitive architecture inspired by the brain.

Cognitive maps are somewhat similar in concept to navigation maps and were proposed back in 1948 [17]. In mammals there is now good experimental proof of cognitive maps allowing spatial navigation [18, 19]. Cognitive maps have been considered in other brain domains and beyond the hippocampal region indirectly [20] and more directly [21]. The Causal Cognitive Architecture considers the long evolutionary importance of navigation maps, defines very specific navigation maps [4], and considers the properties that result from the high-level interactions with each other.

It is important to note that the work this paper reports largely consists of the development of modeling equations [2] and modest Python simulations thereof that can be demonstrated on toy problems. The massive engineering work has not been done to create a robust system of instinctive primitives (discussed below) as well as other features (e.g., even coordinating a myriad of processes occurring in real-time).

Figure 1 illustrates an overview of the Causal Cognitive Architecture 7 (CCA7). As [4] shows, this architecture retains the human-like intelligence properties of its predecessors but now can plan and in particular cases strategize at a superhuman level (albeit, on a conceptual level, given the implementation limitations noted above). The architecture is described in detail and specified formally in [4]. However, to more readily review its operations here, consider the emergence of its properties from its predecessors.

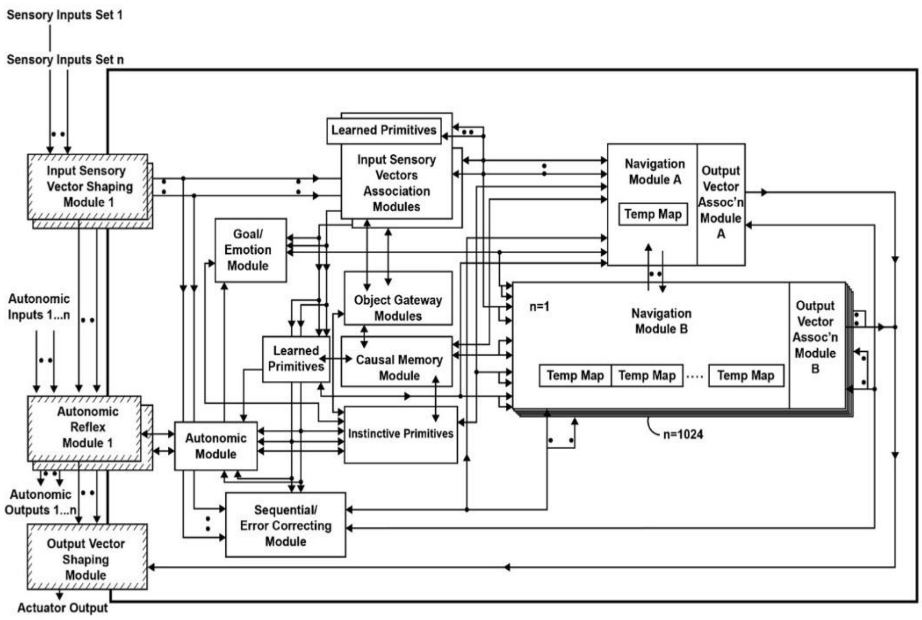


Fig. 1. The Causal Cognitive Architecture 7 (CCA7)

Figure 2 illustrates an overview of the Causal Cognitive Architecture 1 (CCA1) [3]. Sensory inputs stream into the architecture, are pre-processed and normalized (i.e., converted to a common array structure used by the navigation map basic data structure of the architecture) by the Input Sensory Vectors Shaping Module and propagated to the Input Sensory Vectors Association Modules, one module for each different sensory system. Predictive coding essentially occurs here—the input sensory signals are matched to pre-existing navigation maps within each sensory system, which works well for noisy or incomplete sensory inputs. Navigation maps are updated with differences or new navigation maps are created in each of the sensory systems. Binding of the different local sensory navigation maps produced by the input sensory signals with an existing multisensory navigation map from the Causal Memory Module then occurs in the Sensory Vectors Binding Module and is propagated to the Navigation Module. Both spatial binding and temporal binding (via conversion to spatial binding on the navigation map via the Sequential/Error Correcting Module) are further developed in the CCA3 version of the architecture [8]. Note from Figure 2 that the normalized input sensory signal arrays are propagated both to the Input Sensory Vectors Association Modules for eventual spatial binding and the Sequential/Error Correcting Module for binding (i.e., representation on a navigation map with other information) of time-varying signals, i.e., temporal binding.

The processed input sensory signals from the Input Sensory Vectors Association Modules as well as the Sensory Vectors Binding Module will trigger particular Learned Primitives (i.e., learned with experience) or Instinctive Primitives (i.e., pre-programmed) which are essentially small procedures or algorithms which operate on navigation maps. These primitives are typically stored in their respectively named modules but may also be stored in the multisensory navigation maps along with feature data. The best-matching primitive which is triggered is propagated to operate on the multisensory navigation map in the Navigation Module. This may result in an action signal which is transformed into a motor signal by the Output Vector Association Module and Output Vector Shaping Module (Figure 2).

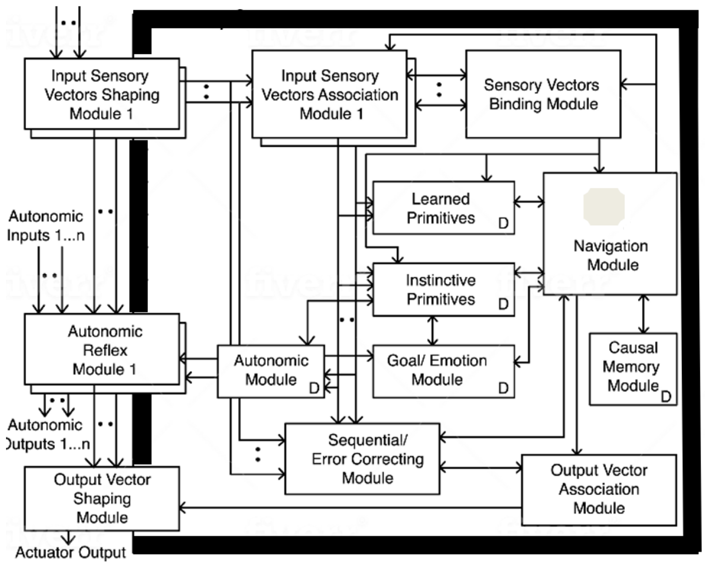


Fig. 2. The Causal Cognitive Architecture 1 (CCA1)

The CCA1 is capable of associative and pre-causal behavior. The navigation maps, i.e., arrays of spatial features and procedures interlinked with other arrays, will generate pre-causal behavior, i.e., not true cause-and-effect but often appears to be due to the storage of features and experiences on the navigation map. However, the predictive coding, i.e., finding differences between input sensory navigation maps and stored existing navigation maps involves large amounts of feedback pathways. If relatively small changes occur (which could also reasonably occur in biological evolution, e.g., [5, 6]) such that feedback is enhanced from the Navigation Module to the Input Sensory Vectors Association, then [3] and [8] show in detail how the intermediate results in the Navigation Module can be fed back and re-processed in the next cognitive cycle (i.e., cycle of sensory inputs flowing through the architecture, although in this case the intermediate results are re-processed rather than new sensory inputs being processed by the Navigation Module), and can be re-processed for a number of cognitive cycles. The effect of re-processing intermediate results allows the architecture to demonstrate full cause-and-effect behavior [3, 8].

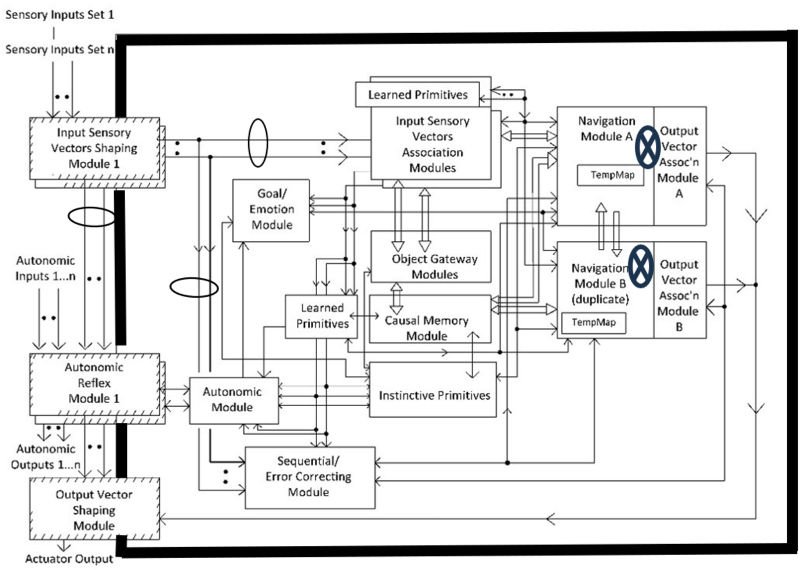
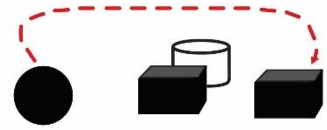


Fig. 3. The Causal Cognitive Architecture 6 (CCA6)

Further small changes in these feedback pathways allow not only full cause-and-effect behavior via re-processing of intermediate results but full analogical reasoning to readily emerge from the architecture, as shown in the CCA5 version of the architecture [9]. Duplication of the Navigation Module (which is biologically possible via modest genetic changes, e.g., [5, 6]) readily allows full compositional properties, including compositional language comprehension to emerge, as shown in the CCA6 version of the architecture (Figure 3) [10]. In Figure 4 the sensory scene shown in 4A is mapped to the Navigation Map shown in 4B which is in Navigation Module A of Figure 3 (or Figure 1). The instruction associated with this sensory scene is mapped to the Navigation Map shown in 4C which is in Navigation Module B of Figure 3 (or Figure 1). The words in the instruction can trigger various instinctive primitives which operate on the Navigation Map in Navigation Module A, and as such, the architecture can compositionally process language. The feedback loops allowing compositional reasoning, including this example, are shown step-by-step in detail in [10].

A



B

air	air	air	air	air	air
air	air	air	air	air	air
air	air	air	air	air	air
air	air	air	air	air	air
air	air	cylinder, white, link {0023,0,0,0}	air	air	air
sphere, black, link {0024,0,0,0}	air	block, black, link {0022,3,3,0}	air	black, block, link {0021,0,0,0}	air

C

"not", link {+}	"near", link {+}	"a", link {+}	"cylinder", link {+}		
"of", link {+}	"the", link {+}	"black", link {+}	"block", link {+}	"which", link {+}	"is", link {+}
"place", link {+}	"the", link {+}	"black", link {+}	"sphere", link {+}	"on", , link {+}	"top", link {+}

Fig. 4. The sensory scene shown in A is mapped to the Navigation Map shown in B which is in Navigation Module A of Figure 3 (or Figure 1). The instruction associated with this sensory scene is mapped to the Navigation Map shown in C which is in Navigation Module B of Figure 3 (or Figure 1).

As a result of a number of modest modifications of an architecture such as the CCA1 (Figure 2) starting from the hypothesis of the neocortex forming from the duplication of navigation circuits in its evolutionary predecessors, the core features of human cognition emerge in the CCA6 (Figure 3)—fully grounded, continuous lifetime learning, associative reasoning, full causal reasoning, analogical reasoning, and near-full compositional language abilities [10].

If further duplication of the Navigation Module B and its temporary memory maps occur along with some additional instinctive primitives (which could hypothetically reasonably occur in biological evolution, e.g., [5, 6]) then the Causal Cognitive Architecture 7 (CCA7) architecture shown in Figure 1 results. As [4] conceptually shows, this architecture in addition to possessing the properties of its predecessor CCA6 architecture, also shows superhuman planning abilities. For example, as [4] shows, there is statistically significant improved performance on a traveling salesperson problem. Note that this occurs from these modest changes to the architecture—there is no hybridization with a large language model (LLM) or other artificial computational module.

2 Modeling of Non-Primate-like AI vs HLAI vs Superhuman AGI vs Alien-like AGI

It is possible to use various combinations of properties or lack of properties of the different versions of the Causal Cognitive Architecture to model "intelligence" or behavior/performance of different artificial systems very roughly based on natural organisms.

In a Python simulation of the Causal Cognitive Architecture, a dropdown menu at the start of the program allows the user the ability to select properties that roughly correspond to a particular group of animals/technology. Table 1 shows which properties of the architecture are used for each particular selection. The Python simulation of the CCA7 version of the architecture is actually used for all natural selections. However, different properties of the architecture may be activated or not in the different selections.

Table 1. Different properties of the CCA7 architecture used for different selections. ("navmaps" stands for "navigation maps")

Animal Group/Tech Selected	**Properties of the CCA7 Architecture Used**
Fish-like brain (i.e., fish-like AI)	reflexive and associative responses, navmaps mainly for actual navigation (illustrated by portions of CCA1 (Fig 2))
Reptilian-like (earlier diapsid-like) brain (i.e., reptilian-like AI)	above + limited number of navmaps with richer associations and aspects of pre-causal behavior emerging (by portions of CCA1 (Fig 2))
Mammalian-like brain (i.e., non-primate mammalian-like AI)	above + larger number of navmaps for all aspects of behavior, more sophisticated instinctive primitives, simple combinatorial language (illustrated by CCA1 (Fig 2))
Human-like brain (i.e., human-level AI (HLAI))	above + enhanced feedback, duplication of navigation modules into A and B, richer instinctive primitives, all resulting in full causality, analogical reasoning and compositional language (CCA6 (Fig 3))
Superhuman-like brain (i.e., superhuman AGI)	above + further duplication of Navigation Module B's, some improved instinctive primitives (illustrated by CCA7 (Fig 1))
Alien-like AGI (ChatGPT3.5 via API)	completely different architecture – large language model -via OpenAI API model "gpt-3.5-turbo" (April 5, 2024) [15]
Alien-like AGI (ChatGPT4) (nb. manual entry)	completely different architecture – large language model -via OpenAI ChatGTP4 (April 5, 2024) [15]

The CCA7 possesses the same properties of its predecessor CCA6 architecture [10] of fully grounded, continuous lifetime learning, associative reasoning, full causal reasoning, analogical reasoning, and near-full compositional language comprehension, but also possesses superhuman planning abilities (albeit, on a conceptual level) and thus is used here as a proxy for a superhuman-like brain or superhuman AGI. Subsets of the CCA7's properties are used as proxies for human-like , mammalian-like, reptile-like and fish-like brains.

For the Alien-like (i.e., non-biologically based) AGI [11], OpenAI's ChatGPT [15] is used as a proxy. The simulation program calls ChatGPT API "gpt-3.5-turbo". A newer ChatGPT4 version was also tested, albeit manually (i.e., outside of the simulation) on the same problems.

3 Method

The models of intelligence in Table 1 need to be evaluated in some fashion. A large number of measures of intelligence exist, including those that are very generalized across a spectrum of animals and machines (e.g., [12, 13]), however, as noted above the largely conceptual CCA7 architecture (which Table 1 utilizes) lacks sufficient instinctive primitives (i.e., preprogrammed) and learned primitives (i.e., learned from experiences) to adequately perform many such intelligence tests. However, the CCA7 does possess a basic planning instinctive primitive and thus can perform planning tests such as the traveling salesperson problem [4]. As well, it does possess a few instinctive primitives that would allow testing of compositionality [10].

The traveling salesperson problem distances for 13 cities are given in [14, 2]. The agent starts at city #0 and must travel to each city once, and then return back to city #0. The goal is to choose a route that minimizes the total distance.

The compositional instruction problem displayed in Figure 4A is utilized along with the instruction to "place the black sphere on top of the black block which is not near a cylinder." Placing the black sphere in the correct location is considered a successful solution; the path taken is not considered.

4 Results

The selections of Table 1, representing different properties of the CCA7 architecture, are applied against the traveling salesperson problem (TSP) and the compositionality problem (CP) described in the previous section. The results are shown in Table 2. The TSP solution can involve randomness at decision points [4]. The CP solution by ChatGPT versions appeared to give random results as well. While compositional problems such the one tested may appear simple to humans as well as to the CCA7 architecture, they are not simple to other primates [10, 16] nor apparently to current large language models.

The simulations of the fish-like brain, reptile-like brain and mammalian-like brain were not able to complete the traveling salesperson problem nor able to correctly follow the compositional instructions.

The p values (Welch's 1-tail t-tests) are computed against the data obtained in the trials of the superhuman-like brain/AGI simulation version.

As noted above, the two versions of OpenAI ChatGPT [15] were used as proxies for an alien AGI [11]. The results from the API version 3.5 were obtained programmatically with zero human intervention. The ChatGPT4 version used manual entry of the problems. For the traveling salesperson problem, within a given trial, ChatGPT4 often attempted and failed to obtain a perfect solution, then tried and failed with another algorithm (with some variation between trials), and then essentially switched to a simple greedy heuristic algorithm yielding a distance of 8131 miles. After the 10th trial, to the prompt was added, "Try this problem again. See if you can use another solution besides the simple, greedy heuristic that you used. It does not have to be a perfect solution." As a result, it yielded better results for the 11th to 20th trials. With regard to the compositionality problem, no additional prompts were provided.

Table 2. Results of attempts to solve the traveling salesperson problem and the compositionality problem (described in the text) by the different selections of the simulation.

Simulated Animal/Tech Group Selected	**Traveling Salesperson Problem**			**Compositionality Problem**		
	n (trials)	**ave distance**	**p (vs superhuman)**	**n (trials)**	**successful**	**p (vs superhuman)**
Fish-like brain/AI	20	all failed	p<0.001	20	0%	p<0.001
Reptilian-like brain/AI	20	all failed	p<0.001	20	0%	p<0.001
Mammalian-like (non-primate) brain/AI	20	all failed	p<0.001	20	0%	p<0.001
Human-like brain/HLAI	20	8131.0	p<0.001	20	100%	--
Superhuman-like brain/AGI	20	7430.2	--	20	100%	--
Alien AGI (ChatGPT 3.5)	20	10221.3	p<0.001	20	3%	p<0.001
Alien AGI (ChatGPT4)	20	7899.6	p<0.001	20	55%	p<0.001

5 Discussion

As noted above, work on the Causal Cognitive Architectures largely consists of the development of modeling equations [2] and modest Python simulations that can be demonstrated on toy problems. The massive engineering work has not been done to create a robust system of instinctive primitives as well as better learning of learned primitives. As a result, a major source of bias in these trials was, of course, choosing problems which the CCA7 at higher architectural levels had been previously tested on. This was done for pragmatic reasons given the limited number of instinctive primitives developed for the architecture and the limited collection of navigation maps of experiences. Nonetheless, the results are still useful to demonstrate a number of issues with regard to pre-mammalian and pre-primate mammalian-like brain functioning/AI and with the development of more advanced artificial intelligence. As well, keep in mind that the CCA7 simulation does not possess a "traveling salesperson" instinctive primitive nor a "compositionality problem" instinctive primitive. Rather, CCA7 simulation contains a simple planning instinctive primitive "`small_plan()`" which is used in planning navigation when there is more than one object to navigate to [4]. (However, as Figure 1 shows, this instinctive primitive is able to run simultaneously in over a thousand Navigation Module B's.) With regard to compositionality, the CCA6 and CCA7 architectures and their simulations have this property intrinsic to these newer architectures essentially [10, 4].

The simulations of the fish-like brain, reptile-like brain and mammalian-like brain were not able to complete the traveling salesperson problem nor able to correctly follow the compositional instructions. The instinctive primitives do not exist that would allow them to successfully complete such a problem. Perhaps in the future there could be better instinctive primitives that allow a reinforcement learning of going to each city once and then returning back to the original city. However, creating such instinctive primitives for the respective architectures of these selections, is far from straightforward. Similarly, these versions of the architecture are not capable of processing compositional instructions and successfully completing the compositional problem. Of interest, other than humans, no mammals, including chimpanzees, are capable of near-full compositional language usage [16]. A dog may exhibit incredibly intelligent behavior but actually getting it to solve the traveling salesperson problem or the compositional problem, let alone training it to solve just those two very specific examples given above, would be highly challenging. Nonetheless, a large variety of tasks that fish-like brains, reptilian-like brains, and mammalian-like brains can accomplish should be included in future simulations so that nonzero scores can be obtained and compared. As well, given the importance of the emergence of causality in the architecture, tasks that better test cause-and-effect abilities are needed.

The human-like brain/HLAI selection gives a traveling salesperson problem distance of 8131 miles, which represents the nearest-neighbor algorithm solution (i.e., just choose the shortest distance at each decision point). However, as noted in [4]. real humans actually have difficulty reliably producing this solution from the numbers alone. The superhuman-like brain/AGI selection, which is based on the full operation of the CCA7 architecture (Figure 1), yields much improved (i.e., shorter) distance for the traveling salesperson problem. As noted above, this version of the architecture can have over a thousand (or as [4] notes, over sixteen-thousand) copies of the planning algorithm running simultaneously.

Future work on the CCA7 architecture simulation includes a larger number of instinctive primitives as well as improvements to the learned primitive system including the storage of a larger set of navigation maps from various experiences.

OpenAI ChatGPT 3.5, as a proxy above for the alien-like AGI, yielded poor results compared to the superhuman brain/AGI selection (Table 2) for both the traveling salesperson problem and the compositionality problem. However, the ChatGPT4 version gave significantly better results in both problems, although again, worse than the superhuman-like brain selection. Given that OpenAI ChatGPT4 is a well-engineered product that has been trained on massive amounts of information compared to the largely conceptual CCA7 simulation, it points to the need to consider intrinsic compositional and planning abilities in the development of AGI systems.

Acknowledgements: Figures 1 and 4 are reproduced under a Creative Commons CC-BY license, with attribution to Schneider, 2024 [4].

References

1. Samsonovich, A.V.: Toward a Unified Catalog of Implemented Cognitive Architectures. In Proc First Annual Meeting of the BICA Society. IOS Press, NLD, 195-244 (2010).
2. Schneider, H.: Supplementary material, (2024). Retrieved May 27, 2024: https://github.com/howard8888/insomnia/blob/master/CCA8_supplementary.pdf
3. Schneider, H.: Causal Cognitive Architecture 1: Integration of connectionist elements into a navigation-based framework. Cognitive Systems Research **66**:67-81 (2021). doi: 10.1016/j.cogsys.2020.10.021
4. Schneider, H.: The Emergence of an Enhanced Intelligence in a Brain-Inspired Cognitive Architecture, Frontiers in Computational Neuroscience, **18**, 1367712 (2024). doi: 10.3389/fncom.2024.1367712
5. Rakic, P.: Evolution of the neocortex: a perspective from developmental biology. Nat Rev Neurosci. Oct;**10**(10):724-35 (2009). doi: 10.1038/nrn2719
6. Chakraborty, M., Jarvis, E.D.: Brain evolution by brain pathway duplication. Philosophical transactions of the Royal Society of London, Series B, **370**(1684), 20150056 (2015).
7. Lieto, A.: Cognitive design for artificial minds. Routledge (2021).
8. Schneider, H.: Causal cognitive architecture 3: A Solution to the binding problem. Cognitive Systems Research **72**:88-115 (2022). doi: 10.1016/j.cogsys.2021.10.004
9. Schneider, H.: An Inductive Analogical Solution to the Grounding Problem. Cognitive Systems Research, **77**:74-216 (2023). doi: 10.1016/j.cogsys.2022.10.005
10. Schneider, H.: The emergence of compositionality in a brain-inspired cognitive architecture. Cognitive Systems Research, **86**, 101215 (2024). doi: 10.1016/j.cogsys.2024.101215
11. Schneider, H., Bołtuć, P.: Alien Versus Natural-Like Artificial General Intelligences. In: Hammer, P., Alirezaie, M., Strannegard, C. (eds) AGI 2023, Lecture Notes in Computer Science, vol 13921:233-243(2023). Springer. doi: 10.1007/978-3-031-33469-6_24
12. Legg, S., Hutter, M.: Universal Intelligence: A Definition of Machine Intelligence. arXiv: 0712.3329 (2007).
13. Chollet, F.: On the measure of intelligence. arXiv preprint arXiv:1911.01547 (2019).
14. Google OR-Tools: Traveling Salesperson Problem (2023). Retrieved Dec 20, 2023 from: https://developers.google.com/optimization/routing/tsp
15. OpenAI: ChatGPT4/API gpt-3.5-turbo (2024). Retrieved 2024 from: https://openai.com/
16. Oña, L. S., Sandler, W., Liebal, K.: A stepping stone to compositionality in chimpanzee communication. PeerJ, **7**, e7623 (2019).
17. Tolman, E. C.: Cognitive maps in rats and men. Psychological review, **55**(4), 189 (1948).
18. O'Keefe, J., Nadel, L. The Hippocampus as a Cognitive Map. Oxford Univ Press, (1978).
19. Alme, C. B., Miao, C., Jezek, K., Treves, A., Moser, E. I., Moser, M. B.: Place cells in the hippocampus: eleven maps for eleven rooms. Proceedings of the National Academy of Sciences of the United States of America, 111(52), 18428–18435 (2014).
20. Behrens, T. E., Muller, T. H., Whittington, et al.: What is a cognitive map? Organizing knowledge for flexible behavior, Neuron, **100**(2), 490-509 (2018).
21. Hawkins, J., Lewis, M., Klukas, M., Purdy, S., Ahmad, S.: A framework for intelligence and cortical function based on grid cells in the neocortex. Frontiers in neural circuits, **12**, 121 (2019).