Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 59 (2020) 73–90

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

The meaningful-based cognitive architecture model of schizophrenia

Howard Schneider

Sheppard Clinic North, Toronto, ON M2M3X4, Canada

Received 1 June 2019; received in revised form 15 September 2019; accepted 15 September 2019
Available online 17 September 2019

Abstract

In subsymbolic operation of the Meaningful-Based Cognitive Architecture (MBCA) the input sensory vector is propagated through a hierarchy of Hopfield-like Network (HLN) functional groups, is recognized and may associatively trigger in the instinctual core goals module as well as in groups of HLNs arranged as pre-causal and pattern memory, vectors propagated to the output motor group of HLNs which produce an output signal. In full causal symbolic operation, the processed sensory input vector is also propagated to the logic/working memory groups of HLNs, where it can be compared to other vectors in the logic/working memory, and produce various outputs in response. The processed sensory input vector can trigger in the instinctual core goals module intuitive logic, intuitive physics, intuitive psychology and intuitive planning procedural vectors, as well as trigger in the causal group of HLNs learned logic, physics, psychology and planning procedural vectors which are also sent to the logic/working memory groups of HLNs. These circuits can allow the MBCA to act causally on information it has never seen before. An example is given of a Python simulation where the MBCA which is controlling a legged robot causally determines that a shallow whitewater river will cause water damage to itself, while if the MBCA is acting associatively only and never having seen whitewater before and normally crossing shallow rivers, will cross the whitewater river and become damaged. While the MBCA does not attempt to replicate biological systems at the neuronal spiking level, its HLNs and the organization of its HLNs are indeed inspired by biological mammalian minicolumns and mammalian brains. The MBCA model leads to the hypothesis that in the course of hominin evolution, HLNs became co-opted into groups of HLNs providing more extensive working memories with causal abilities, unlike non-hominins. While such co-option of the minicolumns can allow advantageous causal symbolic processing integrated with subsymbolic processing, the order of magnitude of increased complexity required for such organization and operation, created a vulnerability in the human brain to psychosis, which does not occur with significant prevalence in non-humans.

Keywords: Cognitive architecture; Artificial general intelligence; Cortical minicolumns; Psychosis; Schizophrenia

1. Introduction

1.1. Achieving human-like causal behavior

Artificial neural networks (ANNs) are capable of pattern recognition and reinforcement learning at a human level of performance (Goodfellow, Bengio, & Courville, 2016; Mnih et al., 2015), but they perform poorly compared to a four-year old child in causally and logically making sense of their environment or the data at hand, particularly when there are only a modest number of training examples (Ullman, 2019; Waismeyer, Meltzoff, & Gopnik, 2015).

Many types of cognitive architectures have already been described in the literature (Langley, 2017; Samsonovich,

E-mail address: howard.schneider@gmail.com

https://doi.org/10.1016/j.cogsys.2019.09.019
1389-0417/

2010). As well, a number of cognitive architectures, for example, ACT-R, CLARION, MicroPsi, Sigma and others, integrate subsymbolic and symbolic processing to varying degrees (Anderson et al., 2004; Bach, 2008; Kilicay-Ergin & Jablokow, 2012; Rosenbloom, Demski, & Ustun, 2016). Research by Graves et al. (2016) uses an ANN which can read and write to an external memory, i.e., a hybrid system. Collier and Beel (2018) note that these newer neural network architectures can be classified as Memory Augmented Neural Networks (MANNs) where there is an external memory unit as opposed to, for example, Long Short Term Memory (LSTM) neural networks where memory exists as an internal vector. MANNs compared to, for example, LSTMs, can perform better on tasks which require memory usage. The Neural Rule Engine of Li, Xu, and Lu (2018) uses modules of neural networks to each represent an action of a logic rule. Work by Albantakis, Hintze, Koch, Adami, and Tononi (2014) simulates the evolution of artificial organisms where the ability to better utilize the causal structure of a rich environment was a driving force for evolution of more complex artificial brains. Lake, Ullman, Tenenbaum, and Gershman (2017) discuss intuitive physics and psychology that exists even in human infants, and the need for machines to build causal models of the world.

Although many of the above architectures and systems can produce both connectionist and symbolic behaviors, they do not fully produce the causal behavior demonstrated by a human child (Waismeyer et al., 2015; Wente et al., 2019). The Meaningful-Based Cognitive Architecture (MBCA) integrates the sensory processing abilities found in artificial neural networks and mammals, with many of the symbolic causal abilities found in human cognition (Schneider, 2018, 2020a).

1.2. Models of schizophrenia

Many types of models, including neural network-based ones, of human schizophrenia and broader psychotic disorders, have already been described in the literature. These range from the work by Cohen and Servan-Schreiber in the 1990s (Cohen & Servan-Schreiber, 1992) to the more recent work of Sabaroedin et al. (2018). As well, there is a large, albeit somewhat speculative, evolutionary psychiatry literature on the origins of schizophrenia. The "schizophrenia paradox" refers to the reality that schizophrenia greatly reduces a person's ability to reproduce, yet it continues to be found throughout the world at a relatively high prevalence of approximately 1%.

Many evolutionary models consider schizophrenic genes as providing an evolutionary advantage and thus evolution balances the risk of schizophrenic genes causing clinical schizophrenia against their purported advantages (Pearlson & Folley, 2008). A wide variety of evolutionary models consider a range of hypothetical physiological, social and behavioral advantages of schizophrenic genes, ranging from the self-domestication (i.e., becoming more tame) of hominids (Benítez-Burraco, Di Pietro, Barba, & Lattanzi, 2017) to schizophrenia allowing some members of prehistoric hunter gatherer tribes to act as shamans who the tribe members believe to have spiritual powers, this being advantageous for the tribes (Polimeni & Reiss, 2003). Work by Crow (2000) hypothesizes that the paradox of schizophrenia persisting in the human population despite reduced fecundity is due to the advantage of humans developing language (a "byproduct" hypothesis of schizophrenia), whereby in schizophrenia there is a failure of adequate brain lateralization associated with language, although recent work by Aase et al. (2018) did not find support for reduced language lateralization.

The original goal of the Meaningful-Based Cognitive Architecture (MBCA) was simply to attempt to produce and experiment with an artificial general intelligence of sorts. The MBCA was never intended to model disease and no disease characteristics were intentionally designed into it. The MBCA does not attempt to replicate biological systems at the neuronal spiking level, but its basic element, a Hopfield-*like* network (HLN), which will be described in more detail below, and the organization of its HLNs, are indeed inspired by biological mammalian minicolumns and mammalian brains (Buxhoeveden & Casanova, 2002; Mountcastle, 1997; Schwalger, Deger, & Gerstner, 2017). As Eliasmith and Trujillo (2014) note, in simulations of brain models perhaps the most important goal is the link to behavior, and indeed at the mesoscopic scale the MBCA can functionally produce a variety of behaviors which can possibly help to better hypothesize and understand mammalian cortical function.

The original informal experimentation with a Python simulation of the MBCA revealed how simple co-option of the basic HLN elements of the MBCA into a pre-causal memory of sorts and then further straightforward co-option into a full symbolic causal processing arrangement, allowed a stream of input data to be processed both subsymbolically (i.e., pattern recognition and associative or reflex behavior) and symbolically (i.e., causal and logical processing of the data in terms of a low level intuitive belief system including intuitive logic, intuitive physics, intuitive psychology and intuitive planning, as well as a similar learned belief system) (Schneider, 2018, 2020a, 2020b). As the informal Python simulation experimentation progressed and the code approached some five thousand lines, an interesting phenomenon rose. While run-time errors and obvious programming errors were caught and were corrected, the informal structure of the code gave rise to numerous design and silent errors which affected mainly the symbolic processing stream. The subsymbolic processing tended to continue to function well—each evaluation cycle (the repeating cycle of the MBCA) after evaluation cycle, the subsymbolic code recognized the input and performed any associative output processing, and then moved onto the next evaluation cycle to process the next input. However, the MBCA's symbolic causal processing using co-opted and modified HLN elementary units was an order of magnitude more complex in requirements for passing data around the MBCA, and retrieval and recognition of the correct data elements. Any of a myriad of code or design issues could cause the symbolic causal processing routines to fail in a fashion almost analogous to biological psychotic behavior.

1.3. Hypotheses

In this paper the following hypotheses are considered:

1. The work below considers qualitatively the hypothesis that the MBCA model approaches the causal behavior seen in humans.
2. The work below considers qualitatively the hypothesis that the emergence of psychotic-like symptoms in the MBCA models the origins of psychotic pathology seen in humans, and offers an explanation with regard to the schizophrenia paradox discussed above.

1.4. Outline

This work will start with an overview of the biologically-inspired but very artificial MBCA. The paper then considers its subsymbolic and symbolic processing in greater detail, examining its causal behavior in a simulated wumpus world environment (Russell & Norvig, 2009, pp. 236–240). The paper then considers the pathological behavior of the *artificial* MBCA and considers the analogous pathological behavior of *biological* mammalian brains, including in particular, human brains.

One is mindful of the large literature encompassing the dozens of subfields the above subject matter falls into. The strategy of Herbert Simon, as annunciated in his 1968 lectures at the Massachusetts Institute of Technology, is followed: "... more we are willing to abstract from the detail of a set of phenomena, the easier it becomes to simulate the phenomena... we do not have to know, or guess at, all the internal structure of the system but only that part of it that is crucial to the abstraction" (Simon, 1996).

2. Overview of Hopfield-like networks (HLNs), meaningfulness and the MBCA

2.1. Hopfield-like network (HLN)

Unlike a typical ANN, the basic unit of the MBCA is not an artificial neuron but a Hopfield-*like* network (HLN). The MBCA uses HLNs for both its subsymbolic (e.g., ANN-like pattern recognition) and its symbolic operations. As well, unlike MANNs, the MBCA does not use an external memory, i.e., it is not a physically hybrid system.

Fig. 1 gives an overview of an HLN unit. An incoming vector (i.e., vector of information representing a sensory input or an input from a previous layer of HLN units) goes to the auto-associative processor. Previous learning experiences have shaped the values of the weights of a given auto-associative processor. If the input vector is recognized, a stored pattern can be outputted to the vector processing unit, and then feeds into the abstraction addressor circuitry. The vector processing unit can be positioned both before and after the auto-associative processor to ensure valid convergence to an output that actually is related to the input vector. Based on the meaningfulness values of the feedback vector (which can be as simple as how many other related HLNs are activated, and will be discussed below), the abstraction addressor decides which of many possible output vectors wired up as inputs to other HLNs will have non-zero outputs. The abstraction addressor effectively allows the HLN to rapidly reconfigure its

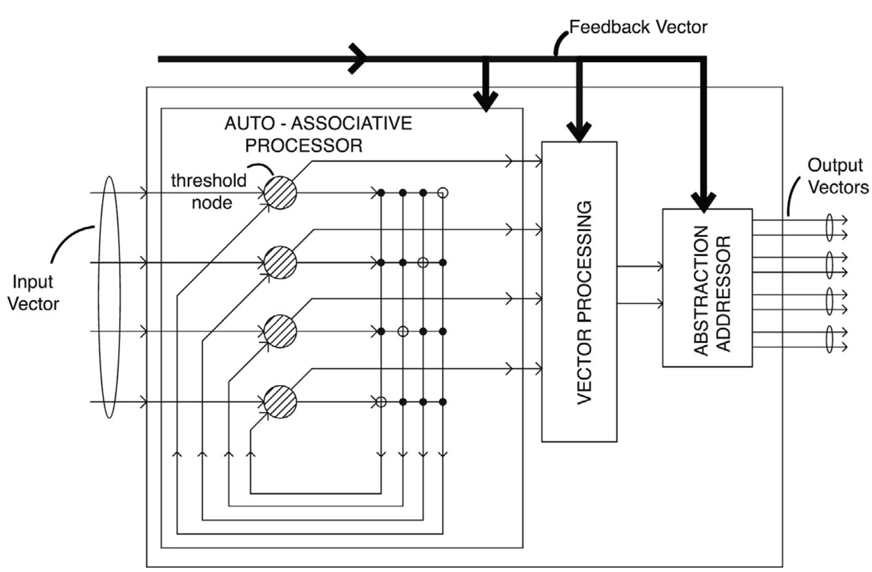

Fig. 1. Overview of a Hopfield-like Network (HLN) unit.

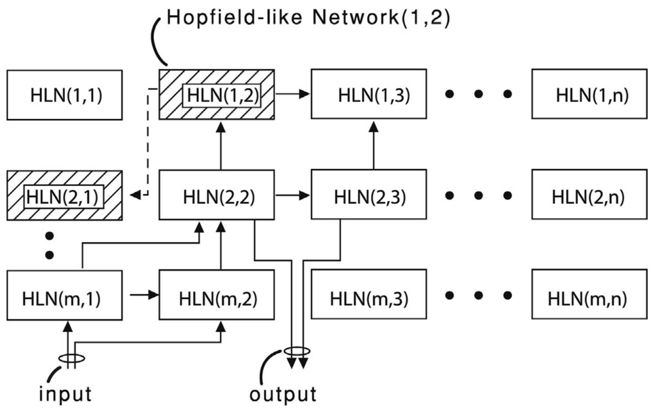

Fig. 2. HLNs are basic units of the MBCA.

connections to other HLNs. (A variety of local algorithms relying on the feedback vector, e.g., an algorithm that tries to keep a high local meaningful value, as well as algorithms that also use computed system meaningfulness values, can decide on the next set of output vectors to activate. The feedback vector can also modulate when the auto-associative processor learns new patterns.)

The MBCA makes use of reconfigurable topologies of Hopfield-like networks (HLNs). Fig. 2, not an organized layer but simply illustrated for demonstrative purposes, shows the output of, for example, HLN(2,3) going to the input of HLN(1,3). HLNs can rapidly reconfigure with other HLNs. In Fig. 2, HLN(1,2) is already connected to HLN(2,1) as shown by the dashed line. Sometimes HLN (1,2) is outputting to HLN(1,3) but other times it is connected to HLN(2,1). Reconfiguration of the HLNs can occur automatically by the HLN units themselves, as well as being influenced by other groups of HLNs and modules of the MBCA.

2.2. Meaningfulness

As noted above, depending on the "meaningfulness" values of the feedback vector, the abstraction addressor decides which of a number of possible sets of output vectors (which become inputs to other HLNs) will have non-zero outputs (Schneider, 2018). The feedback vector can be understood as essentially as asking what reconfigurations are more meaningful in terms of recognizing and processing an input sensory (or intermediate) vector, i.e., essentially as a function of how many other related HLNs are activated. Meaningfulness M is defined as the reciprocal of the Shannon entropy (1,2):

$$H = -\sum_{i} P(x_i) log_2 P(x_i) \quad (1)$$

$$M = 1/H \quad (2)$$

The MBCA is built from reconfigurable topologies of Hopfield-like networks (HLNs). Hawkins and Blakeslee (2004) and Kurzweil (2012) have popularized hierarchies of pattern recognizers inspired by the structure of the mammalian cortex. Hierarchies of pattern recognizers have been discussed in the literature, for example the hierarchical compositional network of George and colleagues (Lázaro-Gredilla, Liu, Phoenix, & George, 2017), and can recognize sensory inputs. However, the construction, operation and many of the properties of the MBCA, from its basic HLN elements to its overall organization is significantly different than in these previous works.

The mechanism of meaningfulness in helping to reconfigure the HLNs to best match an incoming vector is described in more detail and with examples in Schneider (2018). Fig. 3 shows two reconfigurations of the same input sensory HLNs of an MBCA processing an input sensory vector, with the reconfiguration with the higher local meaningfulness value being used to recognize the vector, although in actual implementations meaningfulness is affected by layers higher up as well. Although not proven, it is hoped to extract better sensory signals in noisy environments by such reconfigurations, even in a limited manner, and this can be shown qualitatively. As well, the ability to dynamically reconfigure HLN organization may prove useful in future development of other areas of the MBCA.

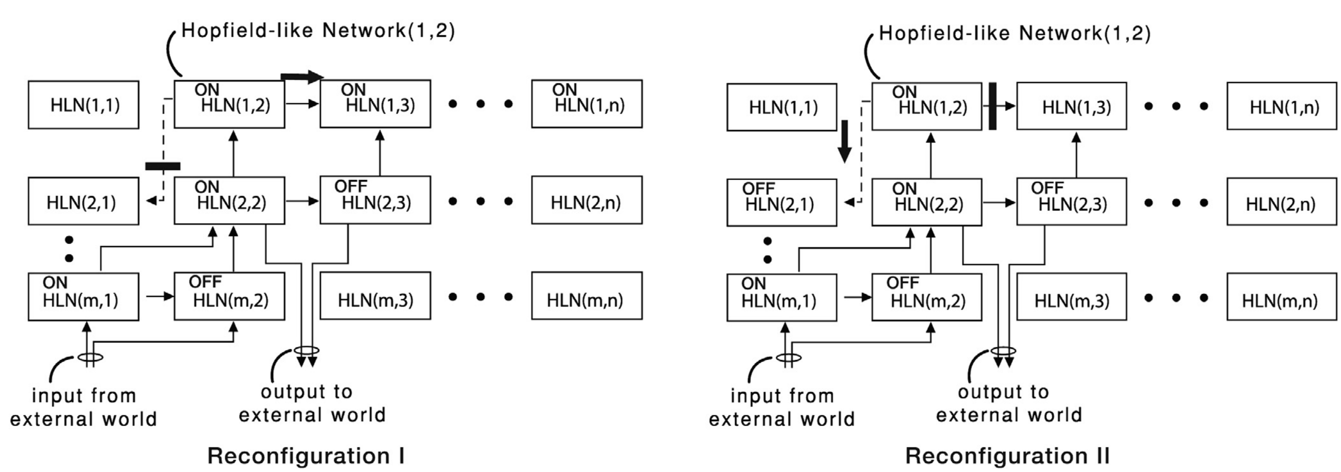

Fig. 3. Same input sensory vector gives Meaningfulness of M = 1.2 (via Shannon entropy of 0.86) for Reconfiguration I versus Meaningfulness = 1.0 (via Shannon entropy of 1.0) for Reconfiguration II.

However, with regard to the main hypotheses raised at the start of this paper, i.e., whether the MBCA models human causal behavior and whether the MBCA models human schizophrenia, the concept of meaningfulness may not be very relevant. Many of the subsymbolic structures of the MBCA, i.e., for learning and matching incoming vectors and triggering various associative behaviors, could actually be replaced with more conventional ANN components. Nonetheless, the HLN was chosen as the basic unit of the MBCA due to the general inspiration provided by the mammalian minicolumn. While a conventional ANN can also pattern match, the HLNs seemed to provide additional useful properties. The ability to try several reconfigurations based on the "meaningfulness" feedback value sometimes gave the ability to better match an incoming vector based on the context of other HLNs, which could be modified by feedback signals from farther layers as well as future symbolic circuits. Further formal analysis as well as more comprehensive simulations are still required to better compare the subsymbolic pattern matching properties of HLNs against various ANNs. As well, using this ability in other non-sensory portions of the MBCA still needs to be better explored. However, of more importance, and of relevance to the hypotheses of causality and of schizophrenia in this paper, is that HLNs seemed to be able to be readily evolutionarily (i.e., straightforward modifications) co-opted into circuits capable of pre-causal and then fully causal symbolic processing, much more so than accomplishing this with the neuron of an ANN.

2.3. *Meaningful-based cognitive architecture (MBCA)*

The Meaningful-Based Cognitive Architecture is shown in Fig. 4. Data from the sensory binding HLNs, the causal memory HLNs, the sequential/error-correcting memory, and other modules and functional units of HLNs, go to the HLNs configured to output motor vectors, which in turn feeds into an output shaping module, and then produces the motor outputs which activate different actuators, as well as providing a data output to the external environment.

If the components in Fig. 4 marked with a solid star are ignored, then the flow of the sensory data and its transformation into motor outputs, is essentially associative and subsymbolic. Consider the "causal memory" in such a configuration to be a "pre-causal memory", as will be shown in more detail below.

Each evaluation cycle (every millisecond or so largely depending on hardware constraints), the input sensory vector is propagated through a hierarchy of reconfigurable Hopfield-like network (HLN) units, and is recognized as some vector (or learned as some new vector). The processed input sensory vector triggers patterns in the "pre-causal memory", in the pattern memory, in the instinctive core goals module and in various other modules and functional groups of HLNs, and these output "procedural vectors" to the pre/causal memory, output motor HLNs, and other areas of the architecture. The procedural vectors essentially cause a series of procedures or actions to occur,

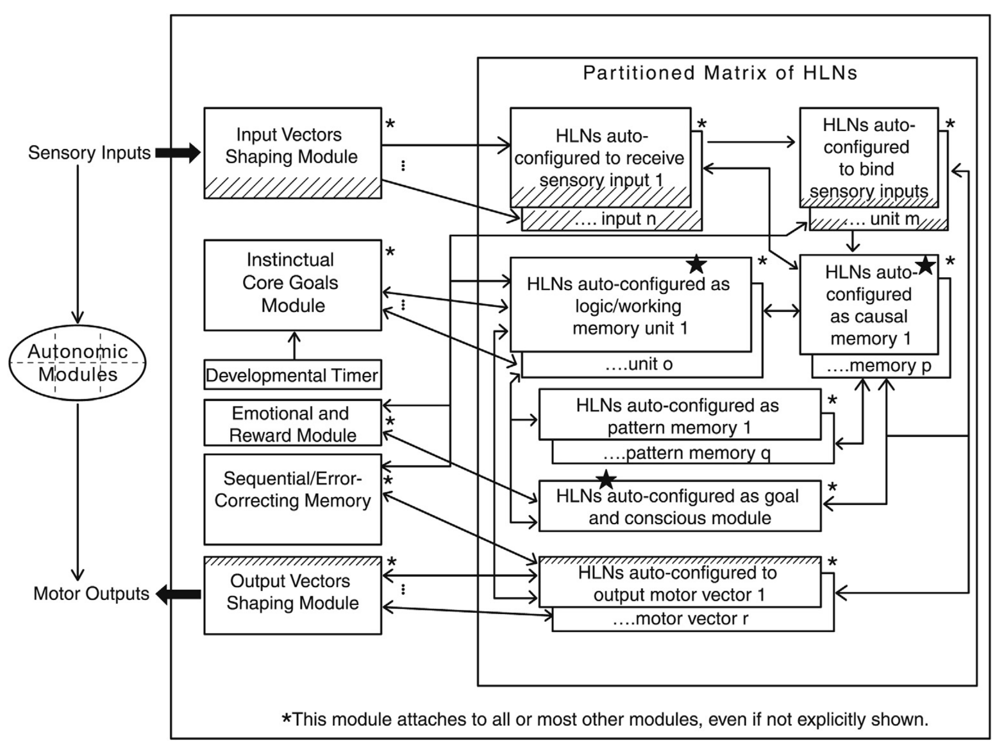

Fig. 4. Meaningful-Based Cognitive Architecture (MBCA)—stars are placed on the symbolic and causal structures.

including possibly causing the output motor HLNs to output a motor vector (i.e., an output signal). There is feedback to the layers of sensory HLNs from the upper levels and from the sensory binding, "pre-causal", pattern and other functional groups of HLNs, and this affects which patterns are recognized. As well, the instinctual core goals module propagates triggered vectors to these regions and will also influence the recognition of the input sensory vector and the triggering of various patterns. Such a system, i.e., *without* the causal and symbolic features of the MBCA discussed below, if tuned well for its environment, can nonetheless be expected to provide a very effective range of complex behaviors. An example of such pre-causal operation of the MBCA is shown below in a simulation.

2.4. Causal symbolic processing of the input sensory data

Groups of HLNs can be co-opted (by design or in a biological system by evolution) away from their original purpose as pattern recognizers to form units dedicated for causal pre-processing of input sensory data. HLNs can easily recognize vectors, and with multiple HLNs, vectors can be compared and selected, and causal processing can be realized.

Groups of HLNs configured as causal memory units are able to:

- match, trigger and receive intuitive procedural vectors from the instinctual core goals module
- match, trigger and receive learned procedural vectors from the local causal memory
- compare properties of vectors they receive
- store a limited number of previous input and output vectors and compare these with the properties of the vectors they receive
- choose one vector over another
- output a vector to the motor output section
- send its output as a feedback signal back to the sensory processing HLNs and modify the processing of the input sensory vector

Each evaluation cycle, the processed sensory input vector is propagated to the instinctual core goals module and to the causal group of HLNs. The processed sensory input vector may trigger in the instinctual core goals module intuitive logic, intuitive physics, intuitive psychology, and/or intuitive goal planning procedural vectors, which are propagated to the causal group of HLNs. Intuitive procedural vectors are the default ones present in a new MBCA, i.e., it is not a tabula rasa. The processed sensory input vector may trigger in the causal group of HLNs learned logic, learned physics, learned psychology, learned goal planning, and/or other learned procedural vectors, which can override the intuitive procedural vectors. The output of the causal group of HLNs is also fed back to earlier sensory processing stages where this feedback will influence the recognition of sensory input features. These groups and modules can send vectors in turn to the output motor group of HLNs which produces an output signal.

The causal group of HLNs thus receives the processed sensory input vector as well as indirectly intuitive and more directly learned procedural vectors. The causal group of HLNs also has temporal information in terms of storage of the one or two previous input and output vectors that propagated to and from the group.

The instinctual core goals module is influenced by the maturity stage of the MBCA via the developmental timer, as shown in Fig. 4. However, the figure does not show the many other pathways affecting the instinctual core goals module. For example, the internal state of the MBCA will also affect the instinctual core goals module's response to vector triggers.

An example of causal operation of the MBCA is shown below in a simulation.

2.5. Simulation of the meaningful-based cognitive architecture

The initial Python code simulation (Schneider, 2018) resulted in the errors which inspired the above hypotheses, i.e., when the causal components were added to the architecture, psychotic-like symptoms (discussed below in detail) emerged in the MBCA for many coding and organizational reasons.

At the time of this writing, the code is being refactored in a more scientifically structured manner such that errors which result in the analogous psychotic behavior of the MBCA can be better specified and actually triggered intentionally for simulation purposes. The "nano" version of the MBCA simulation is the smallest and most reliable version. It does not simulate fully the inner workings of the HLNs or other MBCA structures, but simulates them largely functionally with Python code. Work has started on the "micro" version of the MBCA simulation. It builds out from the "nano" version and there is more actual simulation of detailed aspects of the HLN units and structures of the MBCA.

In the simulation, an implementation of the architecture is controlling a multi-legged search and rescue robot trying to find a lost hiker in an uninhabited forest. The MBCA plus the robot, is named the "MBCA Robot." The MBCA Robot operates inside a simple simulated wumpus world (Russell & Norvig, 2009, pp. 236-240), which contains in addition to the MBCA Robot, the lost hiker, and the environment with areas of forest, lake, shallow rivers and whitewater rapid rivers. The MBCA Robot has a number of simulated sensory transducers and simulated actuators that allow it to move one square north, east, south or west. The goal, of course, is to find the lost hiker, i.e., move to the square where the lost hiker is in.

Fig. 5 shows the initiation of the "nano" version of an MBCA simulation. At the top of the figure the wumpus world set up is shown—a bird's eye view intended for our use, showing where the MBCA Robot is, where the lost

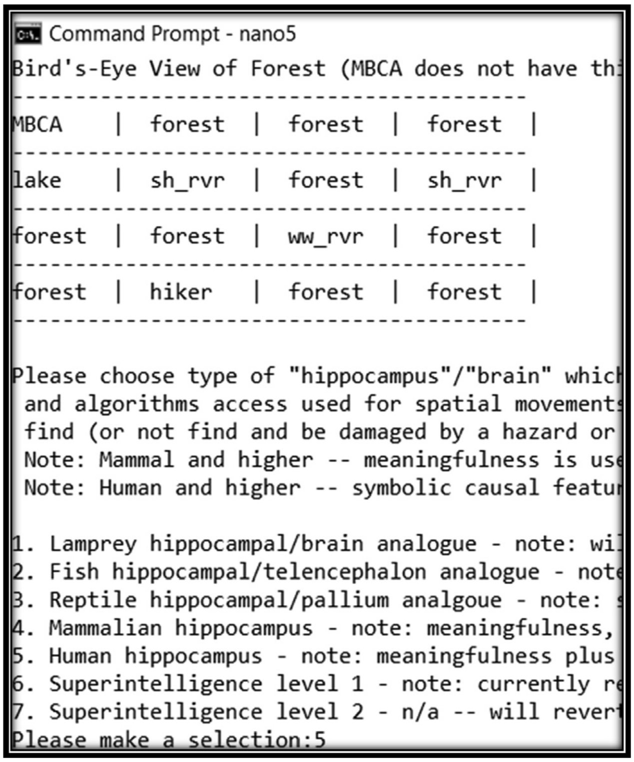

Fig. 5. Initial setup of wumpus world and choice of architecture for MBCA to use.

hiker is, and where the various features of the environment are. The MBCA Robot, of course, does not have this view but must discover the environment on its own, and then construct its own internal map of the world.

In the lower portion of the figure the user is asked to choose a version of the MBCA, ranging from simulation of a lamprey brain to a human hippocampus/brain. (Options for superintelligence simulations are listed as well, but revert back a human brain simulation with certain modules replaced with higher performing ones.) If a choice of "mammalian hippocampus/brain" is made, then the MBCA version will have pre-causal properties and these will be reflected in the resultant simulation. However, if a choice of "human hippocampus/brain" is made, then the version will use the full architecture of the MBCA, and will have causal properties, and these will be reflected in the resultant simulation which is then run.

3. Pre-causal and causal simulations of the MBCA

3.1. MBCA simulation with precausal processing

A simple example of the MBCA Robot operating in a simulated wumpus world will be considered. In the example in this subsection, the MBCA will *not* make use of full causal processing. As discussed above, during initiation of the MBCA simulation, the selection "mammalian hippocampus/brain" is chosen. This engages a version of the MBCA without the main causal features of the architecture.

As described above, the MBCA Robot is trying to find a lost hiker in an uninhabited forest. The MBCA Robot is able to walk through shallow water such as a puddle of water or a shallow river, but if it immerses itself completely in water, for example in a deep lake, it will become damaged. If there is a field or an area of trees, then this area is fine to walk in.

As the MBCA Robot makes its way in searching for the lost hiker, if there is a lake in front of its path, i.e., in front of the direction it wanted to move, it will recognize the lake via its subsymbolic processing (i.e., as shown in Fig. 4, the input sensory vector pathway through a hierarchy of reconfigurable HLN units). This will trigger in the instinctual core goals module an output of "danger/do not walk", and via the output motor group of HLNs, or even as a reflexive action to danger via triggering of the autonomic module, the MBCA Robot will stop. Another direction will be chosen. A bird's eye view of the wumpus world (for the benefit of the observer; the MBCA does not have full knowledge of the wumpus world environment but must construct its own internal world view) is shown in Fig. 6.

On the other hand if there is a shallow river, or a field or a forest, recognized by the MBCA Robot's subsymbolic processing of the visual sensory inputs, the MBCA Robot will continue walking in whatever direction had been chosen. Although causal and higher level symbolic processing can and may be involved in such decisions, they are not required—subsymbolic sensory processing (recognizing the visual sensory input of the lake or shallow river or field or forest) and associative action in the pre-causal memory (i.e., the box labeled causal memory in Fig. 4, but which at this point is really "pre-causal") or elsewhere in the MBCA Robot, will actually suffice.

As the MBCA Robot navigates through the wumpus world, trying to find the lost hiker, it wants to continue south (having hit a number of edges north), and comes in front of a shallow river with many white swirling areas in the river. Subsymbolic processing of the visual inputs

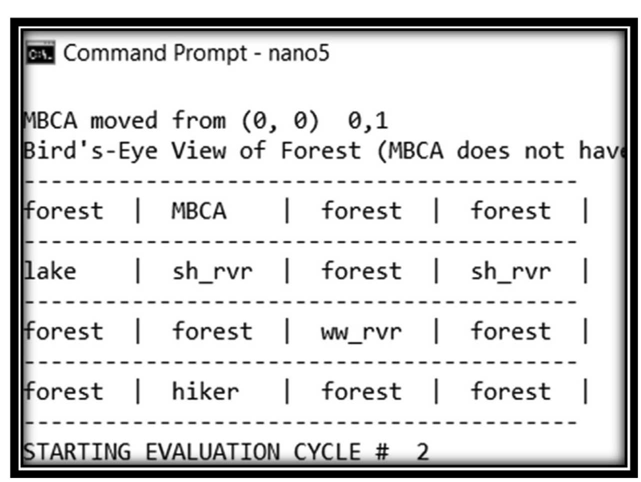

Fig. 6. MBCA Robot Wumpus World Python Simulation—subsymbolic avoidance of a lake.

matches closest to a shallow river (Fig. 7). As noted above, a shallow river is not associated with any danger, and the MBCA Robot can continue its path across it. However, subsymbolic sensory processing also matches smaller areas of the river as white swirling areas. The MBCA Robot has not been programmed with intuitive or learned associative or other actions about what to do if there are white swirling areas in a shallow river. Perhaps another MBCA Robot has tried to cross such a river and the intense pressurized whitewater spray damaged one of that MBCA Robot's leg articulations, and thus after it is retrieved and repaired, the next time that particular MBCA Robot encounters such a river situation it recognizes it as a danger and does not cross it. However, the MBCA Robot in the current example has never seen such a whitewater river before and there is no subsymbolic associative-like action preventing it from crossing the river. Thus it crosses the whitewater river, and unfortunately becomes damaged (Fig. 8). This MBCA Robot is thus unable to rescue the lost hiker, and in fact, must wait for its own rescue.

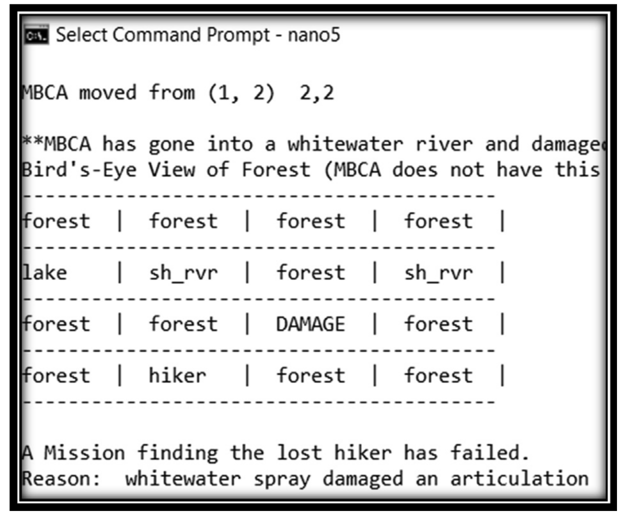

Fig. 8. MBCA Robot Wumpus World Python Simulation—MBCA Robot moves south and is damaged by a high pressure whitewater spray of the river.

3.2. MBCA simulation with full causal processing

A new example of the MBCA Robot trying to find a lost hiker is considered. In this section, during initiation of the MBCA simulation, the selection "human hippocampus/brain" is chosen. This engages a version of the MBCA that uses the full causal features of the architecture.

The causal processing features of the Meaningful-Based Cognitive Architecture, as shown in Fig. 4, are first briefly reviewed here. Every evaluation cycle the MBCA processes the sensory input data. The processed sensory input vector is propagated to the instinctual core goals module and to the causal group of HLNs. The processed sensory input vector may trigger in the instinctual core goals module intuitive logic, intuitive physics, intuitive psychology, and/or intuitive goal planning procedural vectors, which are propagated to the output motor group of HLNs as well as to the causal group of HLNs. The processed sensory input vector may trigger in the causal group of HLNs learned logic, learned psychology, learned goal planning, and/or learned other procedural vectors, which can override the intuitive vectors. The output of the causal group of HLNs is propagated to a number of other groups of HLNs, including the output motor group of HLNs. As well, given the nature of the hierarchy of the HLNs in sensory input processing, the output of the causal group of HLNs is also fed back to earlier sensory processing stages where the feedback signal will influence the recognition of sensory input features. This all happens every single evaluation cycle, over and over again.

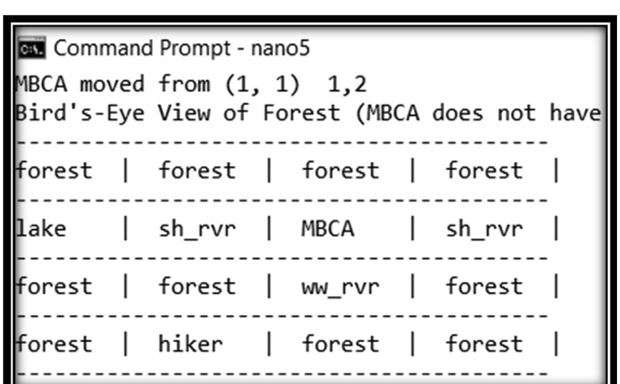

Fig. 7. MBCA Robot Wumpus World Python Simulation—MBCA Robot wants to move south. This (precausal) MBCA Robot has never seen a whitewater river before.

Just as the causal group of HLNs feed back their output to the earlier sensory processing stages, it is possible for the logic/working memory units directly or indirectly via the causal memory units to feed back an intermediate result to the earlier sensory processing stages. In the next evaluation cycle, rather than consider an input sensory vector from the outside world, this intermediate result is propagated through the architecture.

The new example of the MBCA Robot trying to find a lost hiker, this time making use of the full causal features of the architecture, is now considered. Like the "precausal" MBCA Robot in the previous example, this "fully causal" MBCA Robot has never seen or experienced a whitewater river before, nor has it been programmed directly with knowledge about whitewater rivers.

The new MBCA Robot starts off in the wumpus world simulation much like the previous robot did. After a number of moves, it finds itself in front of a shallow river with whitewater currents (Fig. 9). The MBCA Robot has just hit an edge going to the east, and previously hit a number of edges towards the north. As a result intuitive logic procedural vectors have been returned in previous moves directing movement first to west to get away from the edge, and then to the south, i.e., the MBCA Robot's direction preference is to move south to the shallow whitewater river.

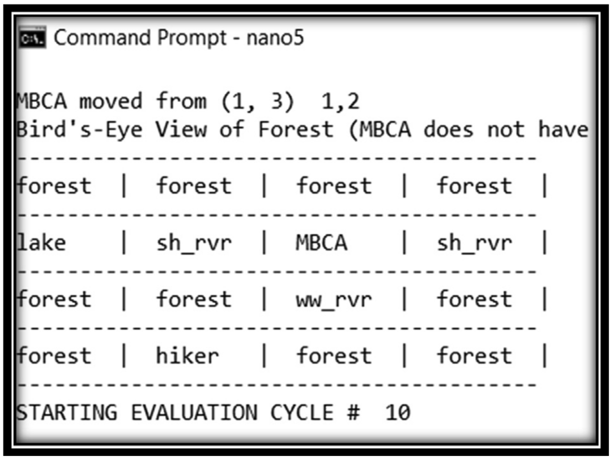

Fig. 9. MBCA Robot Wumpus World Python Simulation— MBCA Robot wants to move south. This (fully causal) MBCA Robot has never seen a whitewater river before.

The southward facing sensory input vectors of "swirling white areas" + "shallow river" are propagated to and trigger in the instinctual core goals module an intuitive physics output of "water all directions," as shown in Fig. 10. The "water all directions" vector is propagated to the logic/working memory and causal memory group of HLNs and is stored there, where it doesn't trigger any action in combination with other values stored there, and thus, the "water all directions" vector is propagated back to earlier sensory stages. The result is that in the next evaluation cycle the processed sensory input vector will be "water all directions" rather than an actual sensory input. The processed sensory input vector—"water all directions"—is propagated, as occurs each evaluation cycle, to the instinctual core goals module where this triggers in its pre-existing intuitive physics knowledge, a "danger/do not walk" vector to the logic/working memory unit, as shown in Fig. 11.

The next evaluation cycle there is processing of the external sensory vector input of the shallow whitewater river, and a processed sensory input vector for "shallow river" arrives at the logic/working memory unit. The logic/working memory unit has the previous procedural vector of "danger/do not walk" stored and now has the processed sensory input vector of "shallow river." The logic/working memory unit sends "shallow river" to the instinctive core goals module where it may trigger the intuitive goal planning current sub-goal of the MBCA Robot, e.g., "walk same direction" and this is propagated back to the logic/working memory group of HLNs. The next evaluation cycle the logic/working memory group of HLNs may ignore the external sensory input vector, but instead send the different procedural vectors, i.e., "danger/do not walk" and "walk same direction" back to the instinctive core goals module. The "danger/do not walk" and "walk same direction" trigger in the instinctive core goals module an intuitive logic procedural vector "change direction ninety degrees." Thus, instead of moving to the south, the MBCA Robot will attempt to move to the east or west, and will avoid the whitewater river in the subsequent moves.

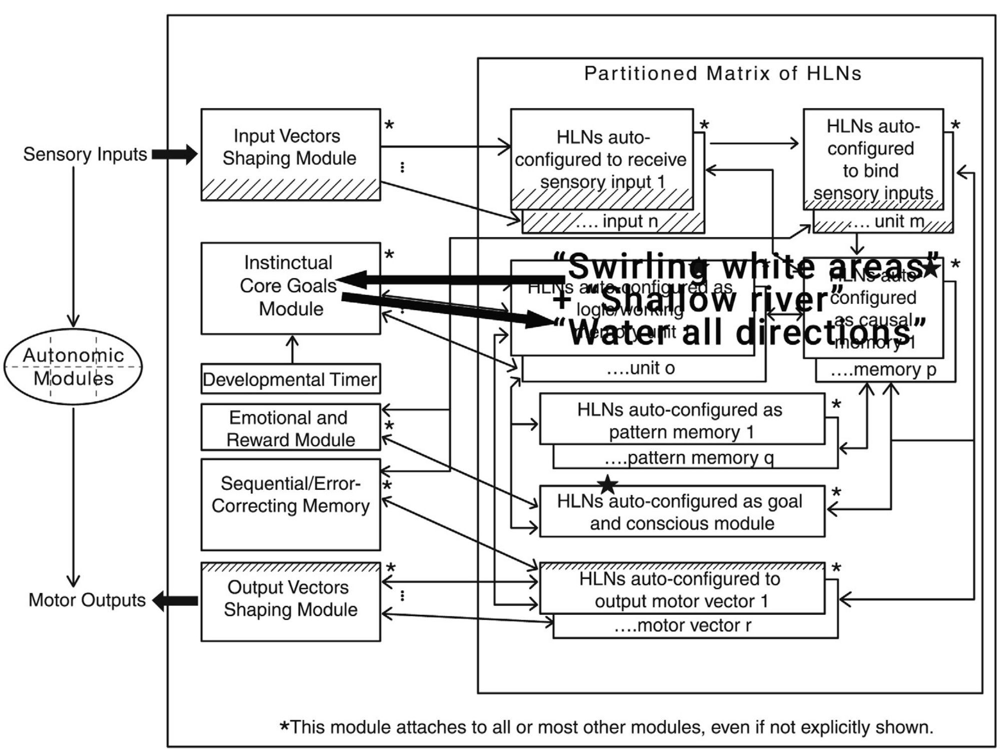

Fig. 10. Meaningful-Based Cognitive Architecture (MBCA)—southward facing sensory input vectors of "swirling white areas" and "shallow river" trigger in intuitive physics (in instinctual core goals module) an output of "water all directions."

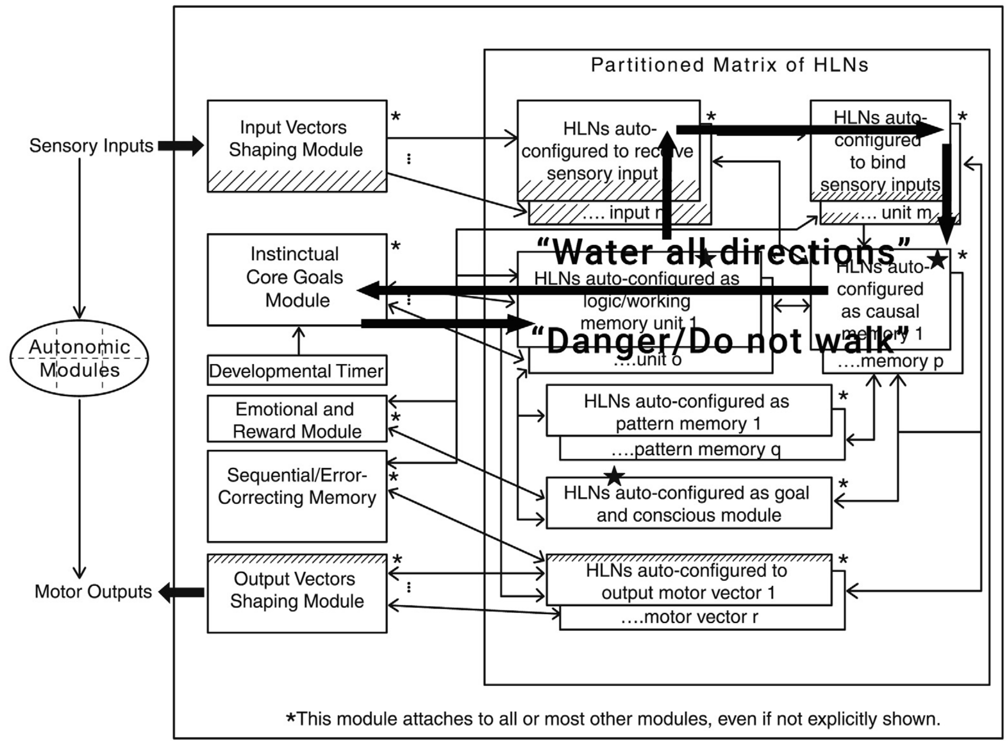

Fig. 11. Meaningful-Based Cognitive Architecture (MBCA)—"water all directions" is propagated back to earlier sensory stages, and in the next evaluation cycle the intermediate result of "water all directions" is treated as the input sensory vector and is propagated to intuitive physics (in the instinctual core goals module) where it triggers "danger/do not walk."

As shown in Fig. 12, after a few more moves the MBCA Robot has indeed gone south, but safely to the side of the whitewater river, not through it. As shown in Fig. 13, after a few more moves, the MBCA Robot has reached and rescued the lost hiker.

This particular MBCA Robot had zero training on recognizing a whitewater river and its properties. Nonetheless, the causal features of the architecture allowed the MBCA Robot to produce causal behavior which was successful in allowing the MBCA Robot to avoid damage from the whitewater river, and successfully rescue the lost hiker.

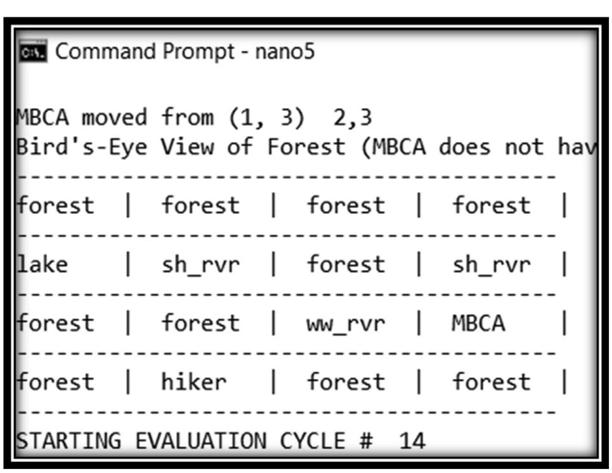

Fig. 12. MBCA Robot Wumpus World Python Simulation—the (fully causal) MBCA Robot has safely moved south.

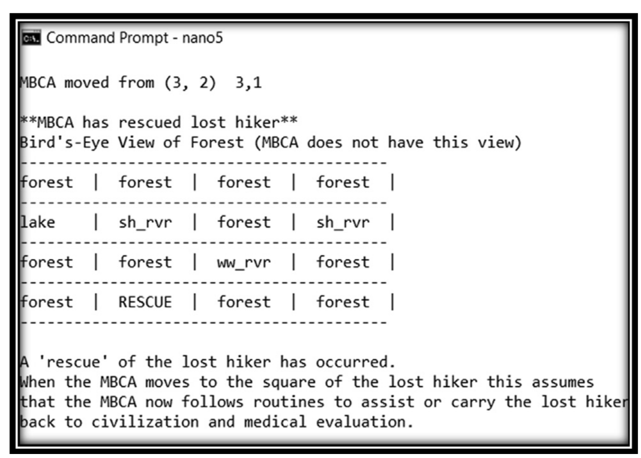

Fig. 13. MBCA Robot Wumpus World Python Simulation—the (fully causal) MBCA Robot has found and rescued the lost hiker, despite never having any previous experience with a whitewater river.

While it may seem that causal behavior should be reserved for special occasions, and perhaps in some other architecture or machine it could be set up as such, the architecture of the MBCA is such that evaluation cycle after evaluation cycle, the processed input sensory vector is also processed causally. Each evaluation cycle, the processed input sensory vector is causally processed against an intuitive and learned belief system, i.e., the intuitive and learned logic, physics, psychology, and planning procedural vectors. Even though some of the time the resultant output behavior ends up being very similar to what an associative mechanism would have outputted, as simulations above reveal, causal processing allows more advantageous behavior in complex environments.

3.3. Further evolution/development and entrenchment of symbolic causal processing from its initial emergence

As noted above, while the MBCA does not attempt to replicate biological systems at the neuronal spiking level, its HLNs and the organization of its HLNs are indeed inspired by biological mammalian minicolumns and mammalian brains (Buxhoeveden & Casanova, 2002; Mountcastle, 1997; Schwalger et al., 2017). While the biological emergence of the logic/working memory units would at first glance seem unfathomable, it becomes much more feasible over modest evolutionary time scales when one considers the development of associative-pseudo-causal groups of HLNs in mammalian brains, and then simple but fuller causal groups, and then with modest evolutionary time, enhancements into full logic/working memory units with a more robust collection of intuitive procedural vectors.

The logic/working memory units can also more readily feed an intermediate result vector back into the sensory input section, and this intermediate result becomes the next processed input sensory vector, as seen in the example above. As such the MBCA can directly work on problems which require numerous sequential logical steps to solve.

Memories of operations occurring in the logic/working memory are kept in the goal and conscious module, shown in Fig. 4. This module is not intended to replicate features of human consciousness, but rather, is simply intended to improve future problem solving by repeating and modifying memories of operations the logic/working memory unit used to solve other problems. However, as a result of this memory of operations, there is good interpretability, albeit for symbolic operations.

The emotional and reward module, shown in Fig. 4, allows effective learning of infrequent events and thus obviates the class imbalance problem seen in conventional neural networks. For example, after the pre-causal MBCA Robot above was damaged in the whitewater river, a memory was made associating shallow rivers with white areas as potentially causing damage. Thus if the damaged MBCA Robot is retrieved and repaired, and then sent back into the forest, when it next sees a whitewater river, even though it has only seen a whitewater river one time before, it will recognize it and will consider it dangerous.

4. MBCA and the emergence of psychotic disorders in *Homo sapiens*

4.1. Emergence of psychotic-like symptoms in the MBCA and humans

As noted above, the original goal of the MBCA was to attempt to produce and experiment with an artificial general intelligence of sorts. The MBCA was not intended to model disease and no disease characteristics were intentionally designed into it, although its HLNs and the organization of its HLNs were indeed inspired by biological mammalian minicolumns and mammalian brains. The original Python simulation of the MBCA was created by informally adding chunks of code and experimenting with the results. As the Python code approached some five thousand lines, silent coding design flaws and actual silent coding errors (i.e., not caught by the interpreter or quasi-unit tests) seemed to affect mainly the symbolic processing stream. The subsymbolic processing was able to make use of a variety of software modules and tools, and tended to continue to function well—each evaluation cycle (the repeating cycle of the MBCA) after evaluation cycle, the subsymbolic code recognized the input and performed any associative output processing, and then moved onto the next evaluation cycle to process the next input. The symbolic causal processing using co-opted and modified HLN elementary units, albeit simulated on some level, was an order of magnitude more complex in requirements for passing data around the MBCA and for retrieval and recognition of the correct data elements. Any one (or multiple) of a myriad of defects could cause the symbolic causal processing routines to fail in a fashion essentially analogous to biological psychotic behavior.

The subsymbolic processing of input sensory vectors is relatively straightforward—each evaluation cycle the MBCA mechanically propagates the input sensory vector through its subsymbolic architecture and arrives at a motor output vector. It is hypothesized that the subsymbolic processing of input sensory vectors by the MBCA is analogous to the operation of the non-hominid mammalian brain—sensory inputs are processed by the subcortical structures as well as the cortex which operates as a pattern matching apparatus. Intuitive logic, intuitive physics, intuitive psychology, and intuitive planning operate both in the subcortical structures as well as the mammalian cortex, and simple pre-causal memories exist to allow simple comparisons. Pattern recognition and associative behavior is taking place, albeit with some pre-causal representation and goal directed influences.

It is hypothesized that in the course of hominin evolution, structures analogous to HLNs became co-opted into groups of HLNs providing more extensive working memories with more logical abilities in comparing vectors, modifying vectors, and outputting vectors, and with more extensive interaction with intuitive and learned logic, physics, psychology, and planning, i.e., the human logic/working memory emerged. As well, the output of the logic/working memory units could be fed back in as a new sensory input, and complex problems could be processed in a number of intermediate steps. The evolutionary advantage of such co-option was more advantageous integrated subsymbolic and symbolic decisions, and the emergence of causal behavior, particularly important to survival in a changing environment.

Unlike the straightforward, mechanical operation of the subsymbolic version of the MBCA, symbolic processing requirements of functionally combining logic/working memory units with the intuitive procedural vectors from the instinctual core goals module and the learned ones from the causal memory, and of shuffling working memory around and retrieving data from the logic/working memory HLNs and HLNs of other functional groups, are an order of magnitude more complex. As the MBCA matures, and as the developmental timer (shown in Fig. 4) causes the instinctual core goals module to make available more complex procedural vectors to the logic/working memory units, any of a myriad of small issues in the logic/working memory units or their connecting modules including the instinctual core goals module, can cause the MBCA to fail in a manner analogous to biological psychosis:

- output vectors from the logic/working memory units are, which themselves may be inappropriate if the output vector arose inappropriately, fed back as a sensory inputs, which the MBCA then inappropriately interprets as a real sensory input rather than a "thought" which a logic/working memory expected to process further in the next evaluation cycle—hallucination-like behavior occurs;
- inappropriately matches and retrieves memory vectors so that they do not correspond with the reality of the input sensory vectors, and inappropriately further processes these vectors – cognitive dysfunction and delusional-like behavior occurs.

Hallucinations, delusions and cognitive dysfunction are the hallmarks of psychosis. There are many medical disorders which involve psychosis, a well-known one being schizophrenia.

4.2. An example of psychotic-like symptoms in the MBCA in a simulated wumpus world

Fig. 14 is the output of a decision move of the MBCA Robot in a Python wumpus world simulation where the MBCA Robot (as above, the term refers to an MBCA controlling a multi-legged robot) is attempting to locate and rescue a lost hiker in an uninhabited forest. The MBCA Robot comes in front of a shallow river which it wants to cross in its search for the hiker. This river has whitewater areas, and even though the MBCA Robot has not encountered whitewater shallow rivers before (and thus has no associative memories about whitewater rivers), the causal processing abilities of the MBCA should allow it to recognize the dangers of a whitewater river, and change direction rather than cross it. This was described in the example above. We would expect that the MBCA feeds the processed sensory input vectors of "swirling white areas" + "shallow river" to the instinctual core goals module where it triggers some sort of intuitive physics response of "water everywhere" which is fed back, processed again, and triggers a danger response and then a switch direction response. However, the fed back "water everywhere" vector somehow triggered an olfactory vector about a smell (a perfume/cologne associated with humans) which could correspond to a lost hiker. However, the smell may not necessarily be real, and in the next step the MBCA Robot goes into the whitewater shallow river, and the pressurized whitewater spray damages its leg articulations. The MBCA

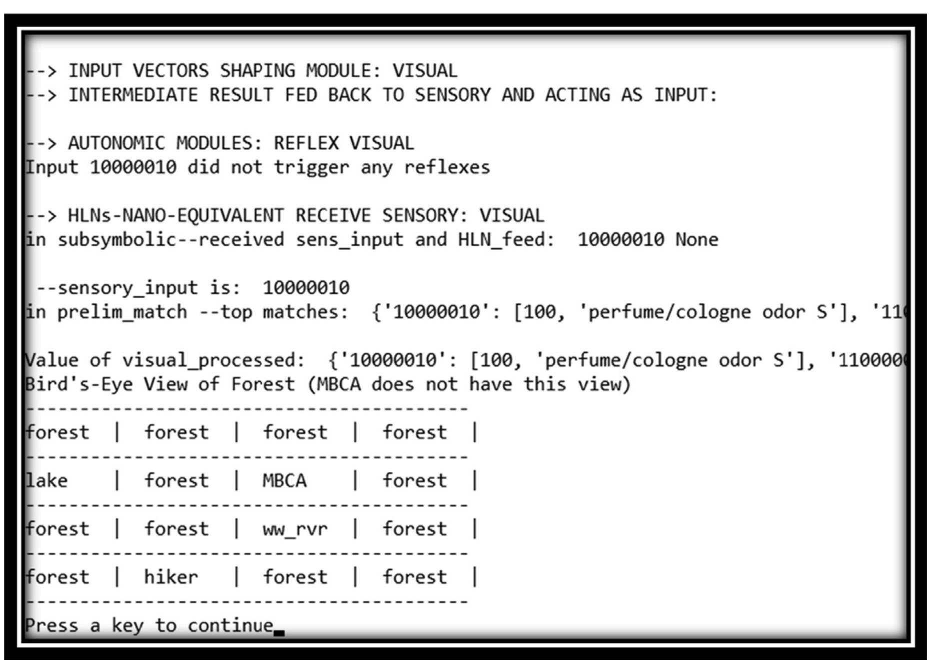

Fig. 14. MBCA wumpus world Python Simulation—symbolic circuitry malfunctioning with inappropriate response to the white water river, and the next move the MBCA moves into the white water river.

Robot now cannot rescue the lost hiker, and itself must now await a rescue.

4.3. MBCA model of human psychotic disorders

As noted above, models of human psychotic disorders, including neural network-based ones, have long been proposed, for example the work by Cohen and Servan-Schreiber in the 1990s (Cohen & Servan-Schreiber, 1992) to the more recent work of Sabaroedin et al. (2018). However, the MBCA was not designed to simulate psychotic disorders, but rather, psychotic features seemed to emerge easily when the co-option of HLNs into logic/working memory had small flaws or there were inconsistencies with the procedural vector from the instinctual core goals module or the causal memory groups.

The MBCA model hypothesizes that an analogous co-option of mammalian cortical minicolumns into effectively parts of logic/working memory units allowed the hominin brain to perform advantageous causal symbolic processing along with subsymbolic processing of input sensory vectors, but the order of magnitude of increased complexity created a vulnerability to psychotic disorders. Indeed, the clinical literature seems to support this hypothesis. Given that the MBCA model predicts that the emergence of psychosis is largely due to a design issue rather than a single defect issue, there should be many reasons why a human could develop psychosis. Indeed, it is found that actually more than 10% of the population (a large figure from a population point of view) will experience less severe psychotic-like symptoms (van Os, Hanssen, Bijl, & Vollebergh, 2001), but just under 1% of the population suffers from schizophrenia (a psychotic disorder)—there are many causes why humans may experience psychotic-like symptoms, as the MBCA model predicts.

In an evaluation of the genomes of 265,218 patients and 784,643 controls it was actually found that there was considerable genetic overlap between what should be very different formal psychiatric disorders including schizophrenia (Anttila et al., 2018), i.e., as the MBCA model predicts psychosis does not emerge from a single or small number of genes.

In just about all other mammals, who would appear by observation to engage in less symbolically processed behavior, psychosis is rare, and in psychopharmacological research settings, large efforts are required to induce at best unreliable models of schizophrenia in research animals (Jones, Watson, & Fone, 2011).

Causal processing by a non-hominin brain should not be assumed, regardless the size of the brain or the ability of the organism to exhibit complex associative behaviors. In a section below, the absence of psychosis and causality in non-humans is considered in more detail, but some of the evidence is reviewed here. Nissani (2006) showed experimentally that in Asian elephants, which have brains much larger than humans, an experimental elephant behavior was due to associative learning rather than due to causal processing. Although by observation it is often reported that (non-mammalian) New Caledonian crows (*Corvus moneduloides*) exhibit causal reasoning in solving physical problems, work by Neilands, Jelbert, Breen, Schiestl, and Taylor (2016) showed that in experiments where the crow must drop a heavy object down a tube in order for food to be released, they do not consider weight nor an understanding of force, and there was little causal understanding. Work by Taylor, Knaebe, and Gray (2012) showed that these crows solve string-pulling problems spontaneously most likely by a perceptual-motor feedback loop rather than insight. Work by Visalberghi and Limongelli (1994) showed that while capuchin monkeys (*Cebus paella*) are able to use a stick to push a reward out of tube, if a trap is added to the tube, the monkeys show little understanding of cause and effect relations in not pushing the food into the trap, and the authors conclude that the monkeys "did not take into account the effects of their actions on the reward." However, it is noted that some mammals may have the beginnings of causal behavior (Sawa, 2009). The tube with a trap and reward-type problem is also difficult for chimpanzees. (The genus *Pan* (common chimpanzee and bonobo species) are the closest extant relatives to humans, and have been typically considered as hominins as well, although there is debate about this classification.) However, work by Seed, Call, Emery, and Clayton (2009) showed that if the tube problem is simplified a bit such that a version of it is presented to the chimpanzees that does not involve the use of a tool, there appears to be some understanding by the chimpanzees about the functional properties of the problem, and they are able to solve it, albeit after dozens of trials. In a section below, the link between causality and psychosis is explored in more detail.

Given that the psychotic behavior in the MBCA emerges from the complexity of co-opting HLNs into logic/working memory units and the operation of these units, and thus above we hypothesize the analogous situation in humans, we would expect that human psychotic disorders would emerge from a large variety of different defects in the working memory circuits or the circuits they attach to. Indeed, lower working memory functioning is found not only in patients with schizophrenia but also in unaffected relatives (Zhang, Picchioni, Allen, & Toulopoulou, 2016).

In young patients at high risk of developing schizophrenia, the medications used to normally treat the symptoms of established schizophrenia do not seem to help with its prevention. However, as the MBCA model predicts, cognitive programs aimed at broadly improving executive function, and thus to some extent working memory, in these high risk young patients, do seem to have a modest effect in preventing the development of schizophrenia (Bechdolf et al., 2012).

4.4. Schizophrenia paradox

As noted earlier, the "schizophrenia paradox" refers to the reality that schizophrenia greatly reduces a person's ability to reproduce, yet it continues to be found throughout the world at a relatively high prevalence of approximately 1%. However, as noted above, the MBCA model of schizophrenia supports that there is not one predominant genetic mutation that causes psychotic symptoms but a whole myriad of them possible due to the evolution of the complex system which allows causal symbolic information processing in the human brain.

The MBCA model rejects the various models of schizophrenia which hypothesize that the genes responsible for schizophrenia or even schizophrenia itself are advantageous on their own in some way, although, of course, over thousands of different genes involved, some variants on their own will have certain advantageous features. Rather, the MBCA model predicts that in a complex, changing environment the ability to perform integrated symbolic/subsymbolic decision making is advantageous compared to a mainly subsymbolic approach, and the vulnerability of the human brain to psychotic disorders is worth the advantage, but where possible without removing the large advantages, this vulnerability will slowly reduce by natural selection.

Work by Liu, Everall, Patnelis, and Bousman (2019) compared the single nucleotide polymorphisms (SNPs) associated with schizophrenia in modern humans with those present in the recovered genomes of extinct archaic Denisovan and Neanderthal humans. The Denisovan and Neanderthal ancestors split from the modern human ancestors approximately 440,000–270,000 years ago (Reich et al., 2010). In comparison the chimpanzee-human last common ancestor is estimated at approximately 8–6 million years ago (although hybridization within the groups may have occurred somewhat more recently than that). It is found that the risk alleles for schizophrenia appear to have been gradually removed from the modern human genome due to negative selection pressure, as the MBCA model would predict.

The MBCA model rejects the schizophrenia paradox. Given that there are many essential genes involved, i.e., the MBCA predicts a design issue rather than a more straightforward genetic disease, the MBCA model would predict that reduction in the gene frequency of at risk alleles would occur very gradually, as the work by Liu and colleagues appears to support.

5. Discussion

5.1. Emergence of causal behavior in humans

The Meaningful-Based Cognitive Architecture (MBCA) was described above. Its architecture is shown in Fig. 4. In subsymbolic operation of the MBCA, an input sensory vector is propagated through a hierarchy of pattern recognizers (i.e., Hopfield-like network (HLN) units). The processed sensory input vector may trigger particular vectors in the instinctual core goals module which itself may send an output vector to the output motor group of HLNs, or back to a pre-causal memory (also composed of variants of HLNs; labeled as “causal memory” in Fig. 4, but is effectively “pre-causal” in subsymbolic operations) which can in turn associatively send an output vector to the output motor group of HLNs. As well, the instinctual core goals module and the pre-causal memory will send a feedback vector to the hierarchy of HLNs processing the sensory input, which will affect how input sensory data will be recognized. The output motor group of HLNs processes the output vectors received and outputs a vector to the output vectors shaping module which in turn outputs a signal to activate actuators and provide a data output.

Above it was discussed how groups of HLNs can be co-opted from their original purpose as pattern recognizers to form units dedicated for causal processing of processed sensory data, all tightly integrated with the subsymbolic structures of the MBCA. These causal groups of HLNs are able to match vectors received and trigger in the instinctual core goals module intuitive logic, intuitive physics, intuitive psychology or intuitive planning procedural vectors which are sent back to the causal groups of HLNs. The causal groups of HLNs can compare properties of vectors they receive and choose one vector over another, and output a vector to the motor HLNs as well as feed back vectors to the hierarchy of sensory processing HLNs.

The MBCA is inspired by an abstraction of the mammalian brain, with the HLNs inspired by mammalian cortical minicolumns. Given that HLNs can be co-opted with a series of small modifications into units dedicated for causal processing of processed sensory data, by analogy it is hypothesized that over a modest amount of evolutionary time, the cortical minicolumns were co-opted into parts of circuits allowing symbolic causal processing integrated with the subsymbolic structures in the hominin brain.

It is hypothesized that natural selection during the evolution of hominins, allowed these causal groups of modified cortical minicolumns to better perform the above operations as well as better match vectors from more of the cortical minicolumns and sub-cortical structures in a hominin brain. As well, it is hypothesized that it became very advantageous during such evolution for the hominin brain equivalent of the instinctual core goals module to provide a larger selection of basic intuitive logic, intuitive physics, intuitive psychology and intuitive planning procedural vectors, plus a better ability to learn and store in the causal memory *learned* logic, physics, psychology and planning procedural vectors. As well, it is hypothesized that the hominin brain also became more capable of processing intermediate results by feeding back these results to the sensory circuits and processing these intermediate results again in the next input cycle. As such it is hypothesized that causal abilities arose in humans, all the while smoothly integrated with associative and reflexive behavior already present.

5.2. Causal behavior and psychosis

The MBCA was not intended to model disease, but as noted above its HLNs and the organization of the HLNs are inspired by mammalian minicolumns and mammalian brains. As the initial Python simulation of the MBCA approached large numbers of lines of code, silent coding and design errors occurred. The subsymbolic processing continued to function well—each evaluation cycle after evaluation cycle, the subsymbolic code recognized the input and performed any associative output processing. However, any of a variety of code issues could cause the symbolic causal processing routines to fail in a way analogous to biological psychotic behavior. The MBCA would occasionally inappropriately retrieve vectors that did not correspond with the reality of the input sensory vectors and then act on these vectors—cognitive dysfunction and delusional-like behavior occurred. The MBCA would occasionally inappropriately interpret intermediate vectors that the instinctual core goals module had fed back to the input sensory hierarchy to process again in the next evaluation cycle, as a real sensory input rather than a "thought" which a logic/working memory expected to process further in the next evaluation cycle—hallucination-like behavior occurred. Hallucinations, delusions and cognitive dysfunction are the hallmarks of psychosis. A number of medical/psychiatric disorders involve psychosis, a well-known one being schizophrenia.

Thus, the MBCA model predicts that in the course of hominin evolution, the biological equivalent of HLNs became co-opted into groups of HLNs providing the ability to process data causally. It is hypothesized that while such co-option of the minicolumns, or other equivalent circuits, can allow functionally valuable causal processing integrated with subsymbolic processing, the order of magnitude of increased complexity required for such organization and operation, created a vulnerability in the human brain to psychotic disorders. The MBCA model predicts that no particular "schizophrenia gene" exists but rather, a vulnerability towards psychosis exists in all humans due to the circuits allowing causal abilities, and a myriad of issues, both genetic and environmental, with these circuits can allow a psychotic state to emerge.

Evidence supporting the MBCA model of schizophrenia and psychosis was given above. In particular:

- MBCA model predicts psychosis including schizophrenia largely stems from a design issue rather than one or a small number of disease genes. Evidence: Anttila et al., 2018—genomes of 265,218 patients and 784,643 controls—considerable genetic overlap between what should be very different formal psychiatric disorders including schizophrenia
- MBCA model predicts psychotic vulnerability in all human brains, thus there should be many reasons why a human could develop psychosis. Evidence: van Os et al., 2001—more than 10% of the population will experience psychotic-like symptoms—even though just under 1% of the population suffers from schizophrenia, a psychotic disorder, there are many other causes why humans may experience psychotic-like symptoms
- MBCA model predicts defects in logic/working memory equivalent in humans related to psychosis. Evidence: Zhang et al., 2016—lower working memory functioning is found not only in patients with schizophrenia but also in unaffected relatives
- MBCA model predicts defects in logic/working memory equivalent in humans are related to psychosis. Evidence: Bechdolf et al., 2012—in young individuals at high risk of developing schizophrenia cognitive programs aimed at broadly improving executive function, and thus to some extent working memory, in these high risk young patients, do seem to have a modest effect in preventing these individuals from developing schizophrenia
- MBCA model rejects the schizophrenia paradox—given that the MBCA model predicts a design issue rather than one or a few genes involved, the MBCA model would predict that the reduction in the gene frequency of at risk alleles would occur very gradually. Evidence: Liu et al., 2019—comparison of SNPs associated with schizophrenia in modern humans with those present in the recovered genomes of extinct archaic Denisovan and Neanderthal humans indicates that risk alleles for schizophrenia appear to be gradually removed from the modern human genome due to negative selection pressure

Given that the MBCA model predicts the emergence of psychotic behavior in humans due to the integration of causal processing with existing reflexive, associative and pre-causal processing existing in early hominins, it raises the question: if other animals do not exhibit true causal behavior, do other animals exhibit psychotic behavior with any frequency?

Evidence supporting the lack of significant causal processing by non-humans was given above. In particular:

- Asian elephants have brains much larger than humans and are anecdotally reported to have human-like behavior. Evidence against: Nissani, 2006—under experimental conditions, elephant behavior found to be due to associative learning rather than the anecdotal appearance of causal processing
- Numerous anecdotal reports of (non-mammalian) New Caledonian crows exhibiting causal problem solving behavior. Evidence against: Taylor et al., 2012—crows solve string-pulling problems spontaneously most likely by a perceptual-motor feedback loop rather than insight

- Numerous anecdotal reports of (non-mammalian) New Caledonian crows exhibiting causal problem solving behavior. Evidence against: Neilands et al., 2016—in experiments where the crow must drop a heavy object down a tube in order for food to be released, they do not consider weight nor an understanding of force, and there was little causal understanding
- Capuchin monkeys appear to be able to use tools anecdotally just like humans. Evidence against: Visalberghi & Limongelli, 1994—while capuchin monkeys are able to use a stick to push a reward out of tube, if a trap is added to the tube, the monkeys show little understanding of cause and effect relations in not pushing the food into the trap, the monkeys "did not take into account the effects of their actions on the reward"
- Chimpanzees are the closest extant relatives to humans, sometimes classified as hominins, and there are anecdotal reports of their use of tools and human-like behavior. Evidence against/equivocal: Seed et al., 2009—like the capuchin monkeys the chimpanzees have trouble understanding cause and effect relations in using a stick to push a reward out of tube if a trap is added, but if the problem is simplified by removing the use of the tool, there appears to be some understanding by the chimpanzees about the functional properties of the problem, and they are able to solve it, albeit after dozens of trials—thus, there may be the beginnings of causal behavior but it is far from the more full causal behavior seen in humans

Do animals exhibit psychotic behavior with any frequency? The answer is generally no. As noted above, the review by Jones et al. (2011) shows that in psychopharmacological research settings, large efforts are actually required to induce at best unreliable models of schizophrenia in research animals.

Thus, this evidence supports that as the MBCA model predicts, the lack of causal processing ability in non-humans appears to be associated with a lack of psychotic behavior of any frequency generally in non-humans.

5.3. Hypothesis of MBCA model approaching the causal behavior seen in humans

At the start of this paper two hypotheses were raised. The first one was whether, from a qualitative point of view, does the MBCA model approach the causal behavior seen in humans, even human children?

Given the obvious toy nature of the wumpus world MBCA simulation compared to an actual human individual, a detailed comparison of behavior is simply not possible. However, what can be gleaned from the wumpus world full MBCA architecture simulation, is a behavior that is indeed qualitatively different from that of a typical ANN. For example, the MBCA Robot had never seen a whitewater river before, yet despite zero training, it was able to recognize what it was to some extent, and it was able to understand its properties, and thus avoid damage by crossing it. This is causal-like behavior, and is qualitatively more typical of humans than typical of ANNs or other non-human biological organisms.

5.4. Hypothesis of MBCA model of schizophrenia being a valid explanation?

At the start of this paper two hypotheses were raised. The second one asked, in a qualitative fashion, whether the emergence of psychotic-like symptoms in the MBCA modelled the origins of psychotic pathology seen in humans, and whether it offered an explanation with regard to the schizophrenia paradox.

Above in Sections 4.3 and 5.2 evidence is given which supports the predictions of the MBCA model of psychosis and schizophrenia.

The MBCA model predicts that in the course of hominin evolution, the biological equivalent of HLNs became co-opted into groups of HLNs providing the ability to process data causally. It is hypothesized that while such co-option of minicolumns, or other equivalent circuits, can allow functionally valuable causal processing integrated with subsymbolic processing, the order of magnitude of increased complexity required for such organization and operation, created a vulnerability in the human brain to psychotic disorders.

As noted above in Sections 4.4 and 5.2, the MBCA model rejects the schizophrenia paradox. The MBCA model predicts a design issue rather than a more straightforward genetic disease. Thus the MBCA predicts that the reduction in the gene frequency of at risk alleles should occur very gradually. The work of Liu et al. (2019), which actually was published a year after the MBCA model was first presented, indeed shows that there is not a schizophrenia paradox but instead risk alleles for schizophrenia appear to be removed very gradually from the modern human genome due to negative selection pressure.

5.5. Future work

Improved MBCA simulations are required that accomplish a number of scientific goals:

- More comprehensive simulations to allow more formal comparison of the subsymbolic pattern matching properties of HLNs with and without meaningfulness, against various ANNs.
- Simulations that can intentionally trigger errors that generate the analogous psychotic behavior of the MBCA, and more formally define the occurrence of these issues in an MBCA.
- Simulations with enhanced intuitive collections of procedural vectors (i.e., more comprehensive collections of intuitive logic, physics, psychology and planning) to

allow simulations of the MBCA in more complex environments to allow the collection of more quantitative performance data.

While the evolutionary aspect of the MBCA simulation, i.e., the ability to simulate an MBCA with features from a lamprey to a human hippocampus/brain, is not a critical feature of the MBCA simulation, it serves to readily illustrate the properties that various features of the architecture contribute to behavior. For example, the behavior of the MBCA in the wumpus world was illustrated with and without the causal features of the architecture. As such, improved non-human models of the MBCA simulations would be helpful.

Pragmatic aspects of the MBCA model include insights into psychotic disorders from a medical point of view, as well as using the MBCA as a guide to the design of systems that will be more capable of artificial general intelligence. Tightly integrating causal symbolic processing with a sub-symbolic artificial neural network may obviate many of the weaknesses of ANNs, including lack of truly understanding the data it processes, poor transparency, and lack of causal abilities.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thanks to D. Kapustin for helpful comments on the manuscript. This article builds upon images originally presented at BICA 2018 (Schneider, 2018).

References

Aase, I., Kompus, K., Gisselgård, J., Joa, I., Johannessen, J. O., & Brønnick, K. (2018). Language lateralization and auditory attention impairment in young adults at ultra-high risk for psychosis: A dichotic listening study. *Frontiers in Psychology, 9*, 608.

Albantakis, L., Hintze, A., Koch, C., Adami, C., & Tononi, G. (2014). Evolution of integrated causal structures in animats exposed to environments of increasing complexity. *PLoS Computational Biology, 10*(12). https://doi.org/10.1371/journal.pcbi.1003966 e1003966.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 111*(4), 1036–1060.

Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., ... Patsopoulos, N. A. (2018). Analysis of shared heritability in common disorders of the brain. *Science, 360*(6395).

Bach, J. (2008). Seven principles of synthetic intelligence. Proceedings of 1st conference on artificial general intelligence, Memphis, pp. 63–74.

Bechdolf, A., Wagner, M., Ruhrmann, S., Harrigan, S., Putzfeld, V., Pukrop, R., ... Bottlender, R. (2012). Preventing progression to first-episode psychosis in early initial prodromal states. *British Journal of Psychiatry, 200*(1), 22–29.

Benítez-Burraco, A., Di Pietro, L., Barba, M., & Lattanzi, W. (2017). Schizophrenia and human self-domestication: An evolutionary linguistics approach. *Brain, Behavior and Evolution, 89*(3), 162–184.

Buxhoeveden, D. P., & Casanova, M. F. (2002). The minicolumn hypothesis in neuroscience. *Brain, 125*(Pt 5), 935–951.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review, 99*(1), 45–77.

Collier, M., & Beel, J. (2018). Implementing neural turing machines. *27th International conference on artificial neural networks (ICANN)*, Rhodes, Greece, October 4–7. .

Crow, T. J. (2000). Schizophrenia as the price that Homo sapiens pays for language: A resolution of the central paradox in the origin of the species. *Brain Research: Brain Research Reviews, 31*(2–3), 118–129.

Eliasmith, C., & Trujillo, O. (2014). The use and abuse of large-scale brain models. *Current Opinion in Neurobiology, 25*, 1–6.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., ... Badia, A. P. (2016). Hybrid computing using a neural network with dynamic external memory. *Nature, 538*, 471–476.

Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. New York, NY: Times Books.

Jones, C. A., Watson, D. J. G., & Fone, K. C. F. (2011). Animal models of schizophrenia. *British Journal of Pharmacology, 164*, 1162–1194.

Kilicay-Ergin, N., & Jablokow, K. (2012). Problem-solving variability in cognitive architectures. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, 42*, 1231–1242.

Kurzweil, R. (2012). *How to create a mind*. New York, NY: Viking Press.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40*, E253. https://doi.org/10.1017/S0140525X16001837.

Langley, P. (2017). Progress and challenges in research on cognitive architectures. Proceedings of the thirty-first AAAI conference on artificial intelligence (AAAI-17).

Lázaro-Gredilla, M., Liu, Y., Phoenix, D. S. & George, D. (2017). Hierarchical compositional feature learning. arXiv preprint arXiv:1611.02252v2.

Li, S., Xu, H., & Lu, Z. (2018). Generalize symbolic knowledge with neural rule engine. arXiv: 1808.10326.

Liu, C., Everall, I., Patnelis, C., & Bousman, C. (2019). Interrogating the evolutionary paradox of schizophrenia: A novel framework and evidence supporting recent negative selection of schizophrenia risk alleles. *Frontiers in Genetics, 10*, 389. https://doi.org/10.3389/fgene.2019.00389.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533. https://doi.org/10.1038/nature14236.

Mountcastle, V. B. (1997). Columnar organization of the neocortex. *Brain, 20*, 701–722.

Neilands, P. D., Jelbert, S. A., Breen, A. J., Schiestl, M., & Taylor, A. H. (2016). How insightful is 'Insight'? New Caledonian crows do not attend to object weight during spontaneous stone dropping. *PLoS ONE, 11*(12). https://doi.org/10.1371/journal.pone.0167419 e0167419.

Nissani, M. (2006). Do Asian elephants (*Elaphas maximus*) apply causal reasoning to tool-use tasks? *Journal of Experimental Psychology: Animal Behavior Processes, 32*(1), 91–96.

Pearlson, G. D., & Folley, B. S. (2008). Schizophrenia, psychiatric genetics, and Darwinian psychiatry: An evolutionary framework. *Schizophrenia Bulletin, 34*(4), 722–733.

Polimeni, J., & Reiss, J. P. (2003). Evolutionary perspective on schizophrenia. *Canadian Journal of Psychiatry, 48*, 34–39.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., ... Maricic, T. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature, 468*(7327), 1053–1060.

Rosenbloom, P., Demski, A., & Ustun, V. (2016). The sigma cognitive architecture and system: Towards functionally elegant grand unification. *Journal of Artificial General Intelligence, 7*(1), 1–103.

Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach* (third ed.). Upper Saddle River, NJ: Pearson.

Sabaroedin, K., Tiego, J., Parkers, L., Sforazzini, F., Finlay, A., Johnson, B., ... Pantelis, C. (2018). Functional connectivity of corticostriatal circuitry and psychosis-like experiences in the general community. *Biological Psychiatry*, pii: S0006-3223(19)30119-2.

Samsonovich, A. V. (2010). Toward a unified catalog of implemented cognitive architectures. *BICA, 221*, 195–244.

Sawa, K. (2009). Predictive behavior and causal learning in animals and humans. *Japanese Psychological Research, 51*(3), 222–233.

Schneider, H. (2018). Meaningful-based cognitive architecture. *Procedia Computer Science BICA 2018, (Ed.) Samsonovich, A.V., 145*, 471–480.

Schneider, H. (2020a). Subsymbolic versus symbolic data flow in the meaningful-based cognitive architecture. In A. V. Samsonovich (Ed.). *Biologically inspired cognitive architectures 2019. BICA 2019. Advances in intelligent systems and computing* (Vol. 948). Springer, Cham. https://doi.org/10.1007/978-3-030-25719-4_61.

Schneider, H. (2020b). Emergence of belief systems and the future of artificial intelligence. In A. V. Samsonovich (Ed.), Biologically inspired cognitive architectures 2019. BICA 2019. Advances in intelligent systems and computing (Vol. 948). Springer, Cham. https://doi.org/10.1007/978-3-030-25719-4_63.

Schwalger, T., Deger, M., & Gerstner, W. (2017). Towards a theory of cortical columns. *PLoS Computational Biology, 13*(4).

Seed, A. M., Call, J., Emery, N. J., & Clayton, N. S. (2009). Chimpanzees solve the trap problem when the confound of tool-use is removed. *Journal of Experimental Psychology: Animal Behavior Processes, 35*(1), 23–34.

Simon, H. A. (1996). *The sciences of the artificial* (third ed.). Cambridge, MA: MIT Press.

Taylor, A. H., Knaebe, B., & Gray, R. D. (2012). An end to insight? New Caledonian crows can spontaneously solve problems without planning their actions. *Proceedings of the Royal Society B: Biological Sciences, 279*, 4977–4981. https://doi.org/10.1098/rspb.2012.1998.

Ullman, S. (2019). Using neuroscience to develop artificial intelligence. *Science, 363*(6428), 692–693.

van Os, J., Hanssen, M., Bijl, R. V., & Vollebergh, W. (2001). Prevalence of psychotic disorder and community level psychotic symptoms: An urban-rural comparison. *Archives of General Psychiatry, 58*(7), 663–668.

Visalberghi, E., & Limongelli, L. (1994). Lack of comprehension of cause-effect relations in tool-using capuchin Monkeys (Cebus paella). *Journal of Comparative Psychology, 108*(1), 15–22.

Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: An action measure. *Developmental Science, 18*(1), 175–182.

Wente, A. O., Kimura, K., Walker, C. M., Banerjee, N., Fernández Flecha, M., MacDonald, B., ... Gopnik, A. (2019). Causal learning across culture and socioeconomic status. *Child Development, 90*, 859–875.

Zhang, R., Picchioni, M., Allen, P., & Toulopoulou, T. (2016). Working memory in unaffected relatives of patients with schizophrenia: A meta-analysis of functional magnetic resonance imaging studies. *Schizophrenia Bulletin, 42*(4), 1068–1077.