



# Neurosymbolic Systems of Perception and Cognition: The Role of Attention

Hugo Latapie<sup>1</sup>, Ozkan Kilic<sup>1\*</sup>, Kristinn R. Thórisson<sup>2</sup>, Pei Wang<sup>3</sup> and Patrick Hammer<sup>4</sup>

<sup>1</sup> Emerging Technologies and Incubation, Cisco Systems, San Jose, CA, United States, <sup>2</sup> Icelandic Institute for Intelligent Machines and Department of Computer Science, Reykjavik University, Reykjavik, Iceland, <sup>3</sup> Department of Computer and Information Sciences, Temple University, Philadelphia, PA, United States, <sup>4</sup> Center for Digital Futures, KTH Royal Institute of Technology and Stockholm University, Stockholm, Sweden

A cognitive architecture aimed at cumulative learning must provide the necessary information and control structures to allow agents to learn incrementally and autonomously from their experience. This involves managing an agent's goals as well as continuously relating sensory information to these in its perception-cognition information processing stack. The more varied the environment of a learning agent is, the more general and flexible must be these mechanisms to handle a wider variety of relevant patterns, tasks, and goal structures. While many researchers agree that information at different levels of abstraction likely differs in its makeup and structure and processing mechanisms, agreement on the particulars of such differences is not generally shared in the research community. A dual processing architecture (often referred to as *System-1* and *System-2*) has been proposed as a model of cognitive processing, and they are often considered as responsible for low- and high-level information, respectively. We posit that cognition is not binary in this way and that knowledge at *any* level of abstraction involves what we refer to as *neurosymbolic* information, meaning that data at both high and low levels must contain *both* symbolic and subsymbolic information. Further, we argue that the main differentiating factor between the processing of high and low levels of data abstraction can be largely attributed to the nature of the involved attention mechanisms. We describe the key arguments behind this view and review relevant evidence from the literature.

**Keywords:** artificial intelligence, cognitive architecture, levels of abstraction, neurosymbolic models, systems of thinking, thalamocortical loop

## OPEN ACCESS

### Edited by:

Alain Morin,  
Mount Royal University, Canada

### Reviewed by:

Mário Boto Ferreira,  
Universidade de Lisboa, Portugal  
Olivier Lionel Georgeon,  
Catholic University of Lyon, France  
Henry Minsky,  
Leela AI, United States

### \*Correspondence:

Ozkan Kilic  
okilic@cisco.com

### Specialty section:

This article was submitted to  
Cognitive Science,  
a section of the journal  
Frontiers in Psychology

**Received:** 31 October 2021

**Accepted:** 06 April 2022

**Published:** 20 May 2022

### Citation:

Latapie H, Kilic O, Thórisson KR,  
Wang P and Hammer P (2022)  
Neurosymbolic Systems of Perception  
and Cognition: The Role of Attention.  
*Front. Psychol.* 13:806397.  
doi: 10.3389/fpsyg.2022.806397

## 1. INTRODUCTION

Cognitive architectures aim to capture the information and control structures necessary to create autonomous learning agents. The sensory modalities of artificially intelligent (AI) agents operating in physical environments must measure relevant information at relatively low levels of detail, commensurate with the agent's intended tasks. Self-supervised learning makes additional requirements on the ability of an agent to dynamically and continuously relate a wide variety of sensory information to high-level goals of tasks. The more general an agent's learning is, the larger a part of its perception-cognition "information stack" must capture the necessary flexibility

to accommodate a wide variety of patterns, plans, tasks, and goal structures. Low levels of cognition (close to the perceptual senses) seem to quickly generate and use predictions to generalize across similar problems. This is a key responsibility of a sensory system because low-latency predictions (i.e., those that the agent can act quickly on) are vital for survival in a rapidly changing world. Natural intelligence has several outstanding skills that Deep Learning does not have. Two of these, as pointed out by e.g., Bengio et al. (2021), are that (a) it does not require thousands of samples to learn, and (b) it can cope with out-of-order (OOD) samples. As detailed by e.g., Thórisson et al. (2019), another equally important shortcoming is that Deep Learning does not handle learning after the system leaves the laboratory—i.e., cumulative learning—in part because it does not harbor any means to verify newly acquired information *autonomously*. Such skills require not only perception processes that categorize the sensory data dynamically so that the lower levels can recognize “familiar” situations by reconfiguring known pieces and trigger higher-level cognition in the case of surprises, but also the reasoning to evaluate the new knowledge that has been thus produced. Whenever high-level cognition solves a new problem, the coordination allows the new knowledge to modify and improve the lower levels for similar future situations, which also means that both systems have access to long-term memory. Architectures addressing both sensory- and planning-levels of cognition are as of yet few and far between.

While general agreement exists in the research community that information at different levels of abstraction likely differs in makeup and structure, agreement on these differences—and thus the particulars of the required architecture and processes involved—is not widely shared. It is sometimes assumed that lower levels of abstraction are subsymbolic<sup>1</sup> and higher levels symbolic, which has led some researchers to the idea that Deep Learning models are analogous to perceptual mechanisms while higher levels involve rule-based reasoning skills due to a symbolic nature, and according to e.g., Kahneman (2011), is the only system that can use language. This view has been adopted in some AI research, where “subsymbolic” processing are classified as System-1 processes, while higher-level and “symbolic” processing is considered belonging to System-2 (c.f. Smolensky, 1988; Sloman, 1996; Strack and Deutsch, 2004; Kahneman, 2011). According to this view, artificial neural networks, including Deep Learning, are System-1 processes; rule-based systems are System-2 processes (see Bengio, 2019; Bengio et al., 2021 for discussion). Similarly, James (1890) proposed that the mind has two mechanisms of thought, one which handled reasoning and another which was associative. We posit instead that cognition is not binary in this way at all, and that *any* level of abstraction involves processes operating on what might be called “*neurosymbolic*” knowledge, meaning that data at both high and low levels must accommodate *both* symbolic and subsymbolic

<sup>1</sup>We classify data as “subsymbolic” if it can only be manipulated through approximate similarity-mapping processes, i.e., cannot be grouped and addressed as a (named) set.

information<sup>2</sup> Further, we argue that a major differentiating factor between the processing of high and low levels of data abstraction can be largely attributed to the nature of the involved *attention mechanisms*.

More than a century ago, James (1890) defined attention as “taking possession by the mind, in clear and vivid form, of one out of what may seem several simultaneously possible objects or trains of thought...It implies withdrawal from some things in order to deal effectively with others.” We consider attention to consist of a (potentially large) set of processes whose role consists in steering the available resources of a cognitive system, from moment to moment, including (but not limited to) its short-term focus, goal pursuit, sensory control, deliberate memorization, memory retrieval, selection of sensory data, and many other subconscious control mechanisms that we can only hypothesize at this point and thus have no names for. Low-level cognition, like perception, is characterized by a relatively high-speed, distributed (“multi-threaded”), subconscious<sup>3</sup> attention control, while higher-level cognition seems more “single-threaded,” and relatively slower. When people introspect, our conscious threads of attention seem to consist primarily of the latter, while much of our low-level perceptions are subconscious and under the control of autonomous attention mechanisms (see Koch and Tsuchiya, 2006; Sumner et al., 2006; Marchetti, 2011 for evidence and discussion about decoupling attention from conscious introspection). Low-level perception and cognitive operations may reflect autonomous access to long-term memory through subconscious attention mechanisms, while higher-level operation may involve the recruitment of deliberate (introspectively-accessible) cognitive control, working memory, and focused attention (Papaioannou et al., 2021).

Two separate issues in the System-1/System-2 discussion are often confused: (1) Knowledge representation and (2) information processing. The first is the (by now, familiar) “symbolic vs. subsymbolic” distinction, while the second involves the “automatic vs. controlled” distinction. Not only are these two distinctly different, they are also not perfectly aligned; while subsymbolic knowledge may be more often processed “automatically” and symbolic knowledge seem generally more accessible through voluntary control and introspection, this mapping cannot be taken as given. A classic example is skill learning like riding a bike, which starts as a controlled process, and gradually becomes automatic with increased training. On the whole this process is largely subsymbolic, with hardly anything but the top-level goals introspectively accessible to the learner of bicycle-riding (“I want to ride this bicycle without falling or crashing into things”). Though we acknowledge the above differences, in this article our focus is on the relations and correlations between these two distinctions. Given the complexity of the project and related experiments mentioned in

<sup>2</sup>By “symbolic” here we mean that the information is at the level of abstraction close to human verbal description, not that it uses “symbols” that must be interpreted or “grounded” to become meaningful.

<sup>3</sup>We consider “subconscious” cognitive processes the set of processes that are necessary for thought and that a mind cannot make the subject of its own cognitive processing, i.e., all its processes that it does not have direct *introspective* access to.

the following sections, this article cannot fully describe the work in detail. It only covers certain aspects of the project within the scope of *attention* suitable for a general audience.

## 2. RELATED WORK AND ATTENTION'S ROLE IN COGNITION

The sharp distinction between two hypothesized systems that some AI researchers have interpreted dual-process theory to entail (cf. Posner, 2020) doesn't seem very convincing when we look at the dependencies between the necessary levels of processing. For instance, it has been demonstrated time and again (cf. Spivey et al., 2013) that expectations created verbally ("System-2 information") have a significant influence on low-level behavior like eye movements ("System-1 information"). It is not obvious why—or how—two sharply separated control systems would be the best—or even a good—way to achieve a tight coupling between levels thus demonstrated, as has been noted by other authors (cf. Houwer, 2019). Until more direct evidence is collected for the hypothesis that there really *are* two systems (as opposed to three, four, fifty, or indeed a continuity), it is a fairly straight forward task to fit the available evidence onto that theory (cf. Strack and Deutsch, 2004). In the context of AI, more direct evidence would include a demonstration of an implemented control scheme that produced some of the same key properties as human cognition from first principles.

We would expect high-level (abstract) and low-level (perceptual/concrete) cognition to work in coordination, not competition, after millions of years of evolution. Rather than implementing a (strict, or semi-strict) pipeline structure between S1 and S2, where only data would go upstream (from S1 to S2) and only control downstream (from S2 to S1; cf. Evans and Elqayam, 2007; Evans and Stanovich, 2013; Keren, 2013; Monteiro and Norman, 2013), we hypothesize high-level and low-level cognition to be coupled through a two-way control-and-data communication, as demonstrated in numerous experiments (see Xu et al., 2020 review article on cross-modal processing between high- and low- level cognition). In other words, the low-level cognition does not solely work under control of the high-level one; rather, the two levels cooperate to optimize resource utilization through joint control.

Through the evolution of the human brain, some evidence seems to indicate that language-based conceptual representations replaced sensory-based compositional concepts, explaining the slower reaction times in humans than other mammals, e.g., chimpanzees (see for instance; Martin et al., 2014). However, this replacement may have pushed the boundaries of human higher-level cognition by allowing complex propositional representations and mental simulations. While animals do not demonstrate propositional properties of human language, researchers have found some recursion in birdsong (Gentner et al., 2006) and in syntax among bonobos (Clay and Zuberbühler, 2011). Moreover, Camp (2009) found evidence that some animals think in compositional representational systems. In other words, animals seem to lack propositional

thought, but they have compositional conceptual thought, which is mostly based on integrated multisensory data. Since animals appear to have symbol-like mental representations, these findings indicate that their lower levels can be neurosymbolic. Evidence for this can be found in a significant number studies from the animal-cognition literature (for review, see Brannon, 2005; Diester and Nieder, 2007; Hauser et al., 2007; Hubbard et al., 2008; Camp, 2009).

Among the processes of key importance in skill learning, to continue with that example, is attention; a major cognitive difference between a skilled bike rider and a learner of bike-riding is what they pay attention to: The knowledgeable rider pays keen attention to the tilt angle and speed of the bicycle, responding by changing the angle of the steering wheel dynamically, in a non-linear relationship. Capable as they may already be of turning the front wheel to any desired angle, a learner is prone to fall over in large part because they don't know what to pay attention to. This is why one of the few obviously useful tips that a teacher of bicycle-riding can give a learner is to "always turn the front wheel in the direction you are falling."

Kahneman (1973) sees attention as a pool of resources which allows different process to share cognitive capabilities and posits a System-1 that is fast, intrinsic, autonomous, emotional, parallel, and a System-2 that is slower, deliberate, conscious, and serial (Kahneman, 2011). For example, driving a car on an empty road (with no unexpected events), recognizing your mother's voice, and calculating  $2+2$ , mostly involve System-1, whereas counting the number of people with eyeglasses in a meeting, recalling and dialing your significant other's phone number, calculating  $13 \times 17$ , and filling out a tax form depend on System-2. Kahneman's System-1 is good at making quick predictions because it constantly models similar situations based on experience. It should be noted that "experience" in this context relates to the process of learning, and its transfer—i.e., generalization and adaptation—which presumably relies heavily on higher-level cognition (and should thus be part of System-2). Learning achieved in conceptual symbolic space can be projected to subsymbolic space. In other words, since symbolic and subsymbolic spaces are in constant interaction, acquired knowledge in symbolic space has correspondences in subsymbolic space. This allows System-1 to start quickly using the projections of the knowledge, even based on System-2 experience.

Several fMRI studies support the idea that sensory-specific areas, such as thalamus, may be involved in multi-sensory stimulus integrations (Miller and D'Esposito, 2005; Noesselt et al., 2007; Werner and Noppeney, 2010), which are symbolic representations in nature. Sensory-specific brain regions are considered to be networks specialized in subsymbolic data that originates from the outside world and different body parts. Thalamo-cortical oscillation is known as a synchronization mechanism or temporal binding between different cortical regions (Llinas, 2002). However, recent evidence shows that the thalamus, previously assumed to be responsible only for relaying sensory impulses from body receptors to the cerebral cortex, can actually integrate these low-level impulses (Tyll et al., 2011; Sampathkumar et al., 2021). In other words, in the thalamus

there are sensory-based integrations, and they are essential in sustaining cortical cognitive functions.

Wolff and Vann (2019) use the term “cognitive thalamus” to describe a gateway to mental representations because recent findings support the idea that thalamocortical and corticothalamic pathways may play complementary but dissociable cognitive roles (see Bolkan et al., 2017; Alcaraz et al., 2018). More specifically, the thalamocortical pathway (the fibers connecting thalamus to cortex region) can create and save task-related representations, not just purely sensory information, and this pathway is essential for updating cortical representations. Similarly, corticothalamic pathways seem to have two major functions: directing cognitive resources (focused attention) and contributing to learning. In a way, the thalamocortical pathway defines the world for the cortex, and the corticothalamic pathway uses attention to tell thalamus what the cortex needs from it to focus. Furthermore, a growing body of evidence shows that the thalamus plays a role in cognitive dysfunction, such as schizophrenia (Anticevic et al., 2014), Down’s syndrome (Perry et al., 2018), drug addiction (Balleine and Leung, 2015), and ADHD (Hua et al., 2021). These discoveries support other recent findings about the role of the thalamus in cognition *via* the thalamocortical loop. The thalamus, a structure proficient in using and integrating subsymbolic data actively, describes the world for the cortex by contributing to the symbolic representations in it. On the other hand, the cortex uses attention to direct resources to refresh its symbolic representations from the subsymbolic space. In Non-Axiomatic Reasoning System (NARS; Wang, 2006) attention has the role of allocating processing power for producing and scheduling inference steps, whereby inferences can compose new representation from existing components, seek out new ones, and update the strength of existing relationships *via* knowledge revision. This control also leads to a refreshing of representations in a certain sense, as the system will utilize the representations which are most reliable and switch to alternatives if some of them turn out to be unreliable.

In the Auto-catalytic Endogenous Reflective Architecture (AERA) attention is implemented as system-permeating control of computational/cognitive resources at very fine-grain levels of processing, bounded by goals at one end and the current situation at the other (cf. Helgason et al., 2013; Nivel et al., 2015). Studies on multitasking in humans have shown that a degree of parallelism among multiple tasks is more likely if the tasks involve different data modalities, such as linguistic and tactile. Low-level attention continuously monitors both mind and the outside world and assesses situations (i.e., relates it to active goals and plans) with little or no effort, through its access to long-term memory and the sensory information. Surprises and threats and detected early in the perceptual stream, while plans and questions are handled at higher levels of abstraction, triggering higher levels of processing, which also provide a top-down control of attention and reasoning. Theoretical foundations and design features including the attention control mechanism of AERA can be found in the detailed technical reports (cf. Thórisson, 2009; Nivel et al., 2013).

In contrast to so-called “attention” mechanisms in artificial neural networks (which are for the most part rather narrow interpretations of resource control in general), mental resources (processing power and storage in computer systems) are explicitly distributed, whereby filtering of input for useful input patterns is just a special case. Another aspect is priming for related information by activating it, which is not limited to currently perceived information but can integrate long-term memory content rather than just content of a sliding window (as in Transformers) of recent stimuli in input space.

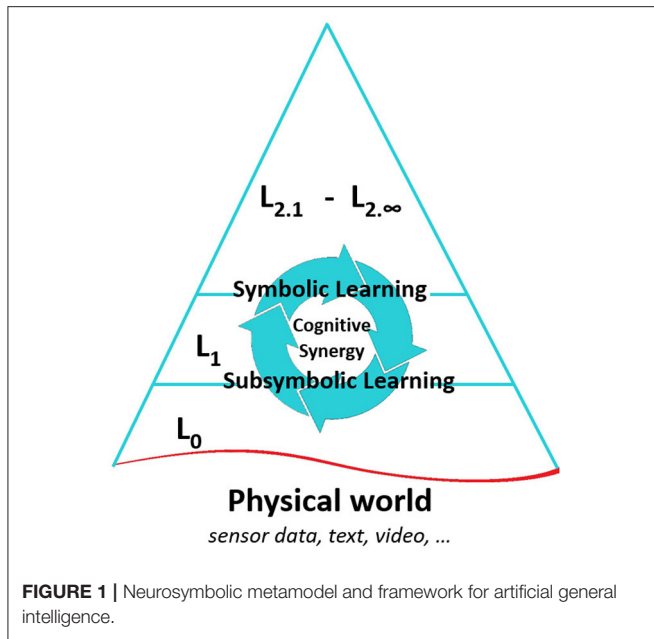
### 3. A NEUROSYMBOLIC ARCHITECTURE AS SYSTEMS OF THINKING

The idea of combining symbolic and sub-symbolic approaches, also known as the neurosymbolic approach, is not new. Many researchers are working on integrated neural-symbolic systems which translate symbolic knowledge into neural networks (or the other way around), because symbols, relations, and rules should have counterparts in the sub-symbolic space. Moreover, the neurosymbolic network needs a symbol manipulation that also supports preservation of the structural relations between the two systems without losing the correspondences.

Currently, Deep Learning and related machine learning methods are primarily subsymbolic. Meanwhile, rule-based systems and related reasoning systems are usually strictly symbolic. We consider it possible to have a Deep Learning model that demonstrates symbolic cognition (without reasoning mechanisms) that entails the transformation of symbolic representations into subsymbolic ML/DL/statistical models. One of the costs associated with such transformation, however, is an inevitable loss of the underlying causal model which may have existed in the symbolic representation (Parisi et al., 2019). Current subsymbolic representations are exclusively correlational; information based on spurious correlation is indistinguishable from other correlations and causal direction between correlating variables is not represented and thus not separable from either of those knowledge sets.

There is an ongoing interest in bringing symbolic and abstract thinking to Deep Learning, which could enable more powerful kinds of learning. Graph neural networks with distinct nodes (Kipf et al., 2018; Steenkiste et al., 2018), transformers with discrete positional elements (Vaswani et al., 2017), and modular models with bandwidth (Goyal and Bengio, 2020) are examples of attempts in this direction. Liu et al. (2021) summarize the advantages of having discrete values (symbols) in a Deep Learning architecture. First, using symbols allows a language for inter-modular interaction and learning, whereby the meaning of symbols is not innate but determined by the relationships with others (as in Semiotics). Second, it allows reusing previously learned symbols in unseen or out-of-order situations, by reinterpreting them in a way suitable to the situation. Discretization in Deep Learning may provide systematic generalization (recombining existing concepts) but it is currently not very successful (Lake and Baroni, 2018).





Current hybrid approaches attempt to combine symbolic and subsymbolic models to compensate for each other's drawbacks. However, the authors believe that there is a need for a metamodel which will accommodate hierarchical knowledge representations. Latapie et al. (2021) proposed such a model inspired by Korzybski's (1994) idea about levels of abstraction. Their model promotes cognitive synergy and metalearning, which refer to the use of different computational techniques and AGI approaches, e.g., probabilistic programming, machine learning/Deep Learning, AERA (Nivel et al., 2013; Thórisson, 2020), NARS<sup>4</sup> (Wang, 2006, 2013) to enrich its knowledge and address combinatorial explosion issues. The current article extends the metamodel as a neurosymbolic architecture<sup>5</sup> as in **Figure 1**.

In this metamodel, the levels of abstractions<sup>6</sup> are marked with L. L0 is the closest to the raw data collected from various sensors. L1 contains the links between raw data and higher level abstractions. L2 corresponds to the highest integrated levels of abstraction learned through statistical learning, reasoning, and other processes. The layer L2 can have an infinite number of sub-layers since any level of abstraction in L2 can have metadata existing at an even higher level of abstraction. L\* holds the high-level goals and motivations, such as self-monitoring, self-adjusting, self-repair, and the like. Similar to the previous version, the neurosymbolic metamodel is based on the assumption of insufficient knowledge and resources (Wang,

<sup>4</sup>With open-source implementation OpenNARS at <https://github.com/opennars/opennars> (accessed October 20th, 2021).

<sup>5</sup>This architecture has a "symbolic" aspect in the sense that there are components that can be accessed and manipulated using their identifiers. This is different from traditional Symbolic AI where a "symbol" gets its meaning by referring to an external object or event, as stated by Newell and Simon (1976).

<sup>6</sup>Korzybski (1994) states that knowledge is a multiordinal, hierarchical structure with varying levels of abstraction.

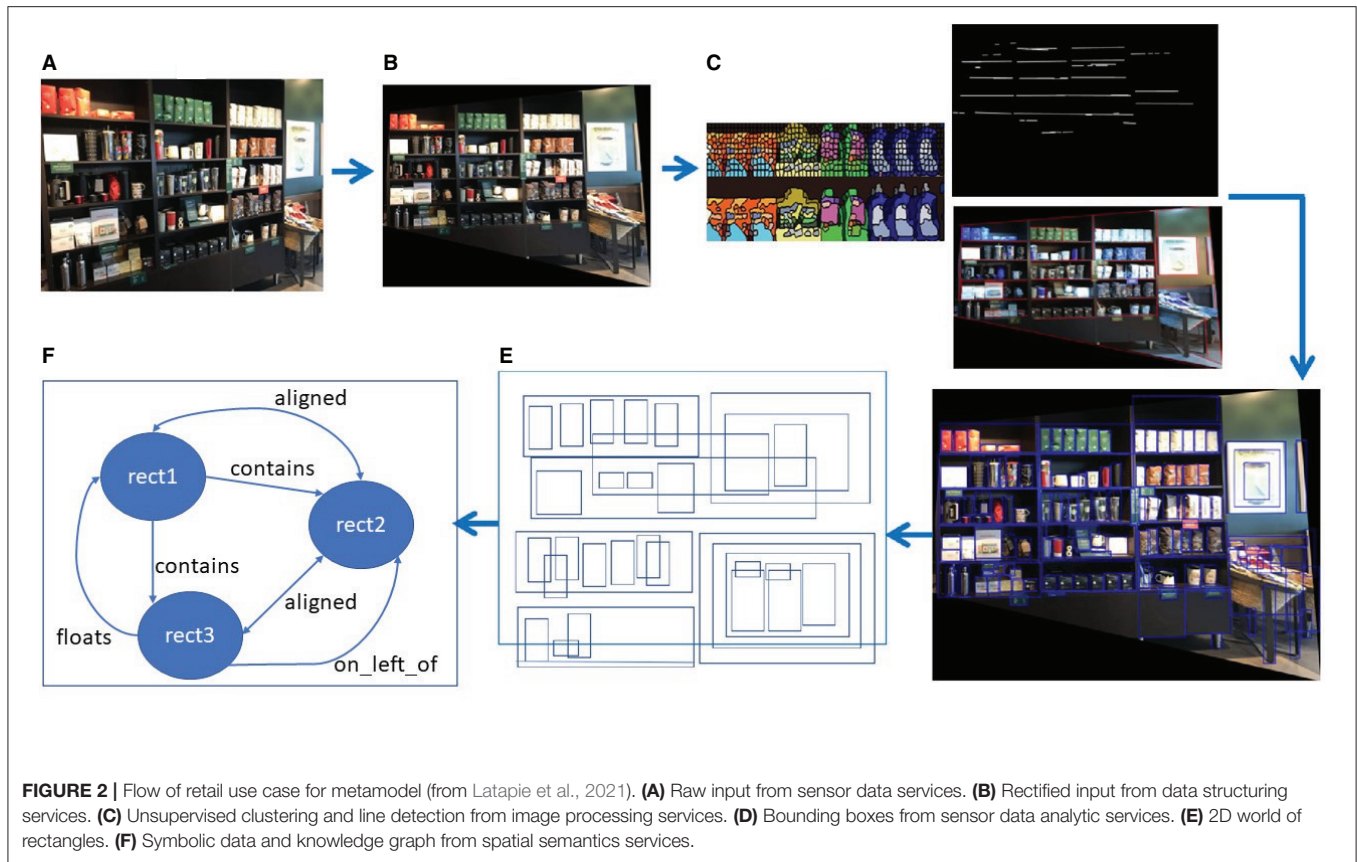
2005). The symbolic piece of the metamodel can be thought of as a knowledge graph with some additional structure that includes both a formalized means of handling anti-symmetric and symmetric relations, as well as a model of abstraction. The regions in the subsymbolic piece of the metamodel are mapped to the nodes in the symbolic system in L1. In this approach, the symbolic representations are always refreshed in a bottom-up manner.

Depending on the system's goal or subgoals, the metamodel can be readily partitioned into subgraphs using the hierarchical abstraction substructure associated with the current focus of attention. This partitioning mechanism is crucial to manage combinatorial explosion issues while enabling multiple reasoners to operate in parallel. Each partition can trigger a sub-focus of attention (sFoA), which requests subsymbolic data from System-1 or some answers from System-2. The bottom-up refreshing and the neurosymbolic mapping between regions and symbols allow the metamodel to benefit from different computational techniques (e.g., probabilistic programming, Machine Learning/Deep Learning and such) to enrich its knowledge and benefit from the 'blessing of dimensionality' (cf. Gorban and Tyukin, 2018), also referred to as "cognitive synergy."

A precursor to the metamodel as a neurosymbolic approach was first used by Hammer et al. (2019). This version was the first commercial implementation of a neurosymbolic AGI-aspiring<sup>7</sup> approach in the smart city domain. Later, the need for use of the levels of abstraction in the metamodel became mandatory due to the combinatorial explosion issue. In other words, structural knowledge representation with the levels of abstraction became very important for partitioning the problem, process subsymbolic or symbolic information for each sub problem (focus of attention, FoA), and then combine the symbolic results in the metamodel. The metamodel with the level of abstraction was actually achieved fully in the retail domain (see Latapie et al., 2021 for details). The flow of the retail use case with the metamodel is shown in **Figure 2**. The example for the levels of abstraction using the results of the retail use case is shown in **Figure 3**. Latapie et al. (2021) emphasized that no Deep Learning model was trained with product or shelf images for the retail use case. The system used for the retail use case is solely based on representing the subsymbolic information in a world of bounding boxes with spatial semantics. The authors tested the metamodel in four different settings with and without the FoA and reported the results as in **Table 1**.

Another use case for the metamodel is the processing of more than 200,000 time series with a total of more than 30 million individual data points. The time series are network telemetry data. For this use case, there are only two underlying assumptions: The first assumption is that the time series or a subset of them is at least weakly-related, such as time series from computer network devices. The second assumption is that when a number of time series simultaneously change their behaviors, it might indicate that an event-of-interest has happened. For

<sup>7</sup>Artificial general intelligence (AGI) is the research area closest to the original vision of the field of AI, namely, to create machines with intelligence on par with humans.



**FIGURE 2 |** Flow of retail use case for metamodel (from Latapie et al., 2021). **(A)** Raw input from sensor data services. **(B)** Rectified input from data structuring services. **(C)** Unsupervised clustering and line detection from image processing services. **(D)** Bounding boxes from sensor data analytic services. **(E)** 2D world of rectangles. **(F)** Symbolic data and knowledge graph from spatial semantics services.

detecting anomalies and finding regime change locations, Matrix Profile algorithms are used (see Yeh et al., 2016; Gharghabi et al., 2017 for Matrix Profile and Semantic Segmentation). Similar to the retail use case, millions of sensory data points are reduced to a much smaller number of events based on the semantic segmentation points. These points are used to form a histogram of regime changes as shown in **Figure 4**. The large *spikes* in the histogram are identified as the *candidate events-of-interest*. Then the metamodel creates a descriptive model for all time series, which allows system to downsize millions of data points into a few thousand structural actionable and explainable knowledge.

To test the metamodel with time series, we first use a subset of the Cisco Open Telemetry Data Set<sup>8</sup>. After being able to identify the anomalies in the data set, we create our own data sets similar to the Open Telemetry Data. For this purpose, 30 computer network events, such as memory leak, transceiver pull, port flap, port shut down, and such, are injected to a physical computer network. The system is able to identify 100% of the events with a maximum of 1 minute delay. For example, **Figure 4** represents the histogram of regime changes for a *port shut down* event, which is injected at the 50th timestamp. Since the sampling rate is 6 s, 1 min later (which is at the 60th timestamp) the system detects a spike as an event-of-interest. It can take time for a single incident to display a cascading effect on multiple devices. When the injection ends at the 100th

timestamp, another spike is observed within 10 timestamps, which represents a recovery behavior for the network. It should be noted that not all events necessarily mean an error has happened. Some usual activities in the network, e.g., a usual firmware update on multiple devices as events-of-no-interest, are also captured by the metamodel. The metamodel learns to classify such activities either by observing the network. Although the time series processing using the metamodel does not require any knowledge of computer networking, it can easily incorporate such features extracted by networking-specific modules, e.g., Cisco Joy,<sup>9</sup> or ingest some expert knowledge defined in the symbolic world, specifically at the 2nd level of abstraction This neurosymbolic approach with the metamodel can quickly reduce the sensory data into knowledge, reason on this knowledge, and notify the network operators for remediation or trigger a self-healing protocol.

## 4. DISCUSSION

The neurosymbolic approach presented here evolved from several independent research efforts by four core teams [NARS, AERA, OpenCog (Hart and Goertzel, 2008)], all of which are open source projects) as well as efforts at Cisco over the past 10 years focusing on hybrid state-of-the-art AI for commercial applications. This empirically-based approach to AI

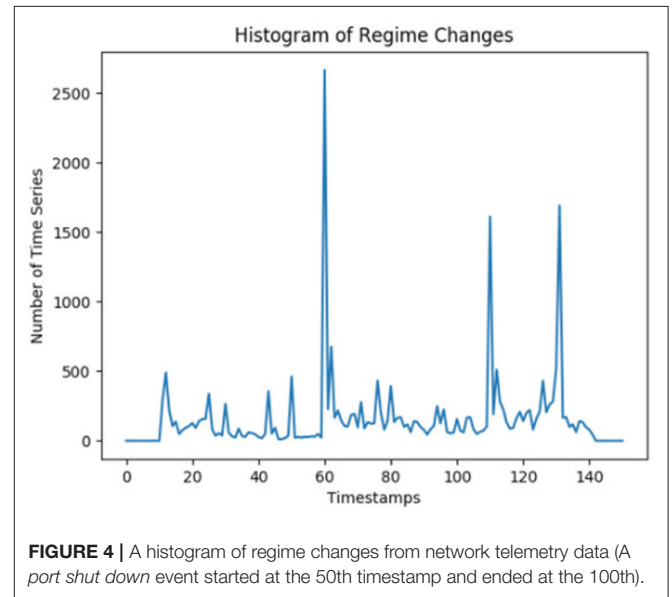
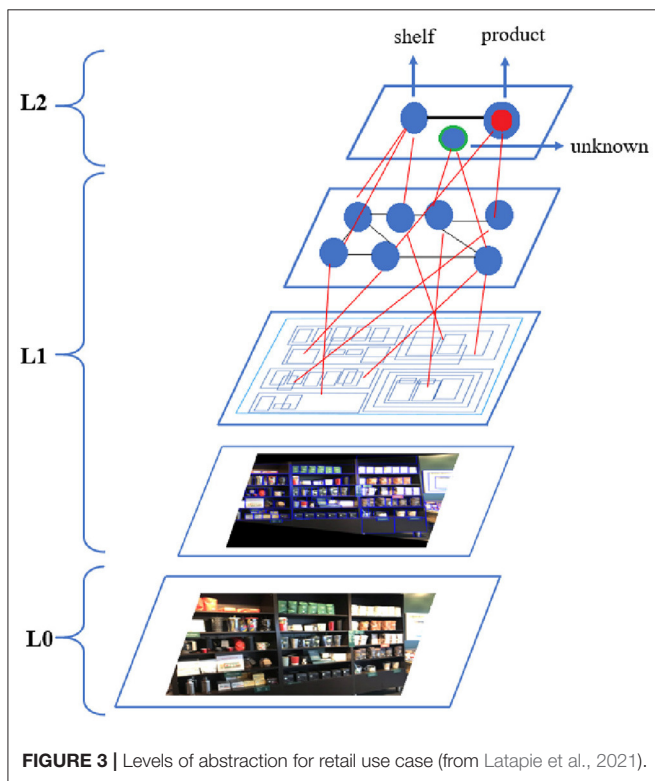
<sup>8</sup><https://github.com/cisco-ie/telemetry>

<sup>9</sup><https://github.com/cisco/joy>

took off (circa 2010) with deep-learning based computer vision, augmented by well-known tracking algorithms (e.g., Kalman filtering/Hungarian algorithm). The initial hybrid architecture resulted in improved object detection and tracking functionality, but the types of errors, arguably related to weak knowledge representation and poor ability to define and learn complex behaviors, resulted in systems which did not meet our performance objectives. This initial hybrid architecture was called DFRE, Deep Fusion Reasoning Engine, which actually lacked the metamodel. In order to improve the system’s ability to generalize, NARS was incorporated. The initial architecture used NARS to reason about objects and their movements in busy city intersection with trains, busses, pedestrians, and heavy traffic. This initial attempt at a commercial neurosymbolic system dramatically improved the ability of the system to generalize and learn behaviors of interest, which in this case were all related to safety. In essence the objective of the system was to raise alerts if any two moving objects either made contact

or were predicted to make contact as well as to learn other dangerous behaviors such as jay walking, wrong-way driving, and such. While this system worked well as an initial prototype and is considered a success, there were early indications of potential computational scalability issues if the number of objects requiring real-time processing were to increase from the average 100 or so to say an order of magnitude more objects, such as 1,000. In order to explore this problem we then focused on a retail inventory use case that required the processing of over 1,000 objects. As expected, DFRE suffered from the predicted combinatorial explosion issues. In the retail use case, this problem was solved *via* the metamodel’s abstraction hierarchy which provides a natural knowledge partitioning mechanism. This partitioning mechanism was used to address the exponential time complexity problem and convert it to a linear time complexity problem.

While NARS enabled the system to learn by reasoning in an unsupervised manner, there was a growing need in commercial applications for a principled mechanism for unsupervised learning directly from temporal data streams such as sensor data, video data, telemetry data, etc. This is the focus of AERA as well as internal Cisco project Kronos based on Matrix Profile (Yeh et al., 2016). While there is a large body of work on time series processing (FFT, Wavelets, Matrix Profile, etc.), the



**TABLE 1 |** Experimental results from retail use case using metamodel.

| Category         | Without FoA (%)                     |        |          | With FoA (%)                         |        |          |
|------------------|-------------------------------------|--------|----------|--------------------------------------|--------|----------|
|                  | Precision                           | Recall | f1-score | Precision                            | Recall | f1-score |
| Product          | 80.70                               | 29.32  | 52.88    | 96.36                                | 99.07  | 97.70    |
| Shelf            | 8.82                                | 18.75  | 12.00    | 82.35                                | 87.50  | 88.85    |
| Other            | 36.61                               | 89.66  | 52.00    | 96.00                                | 82.76  | 88.89    |
| Overall accuracy | <b>46.30</b> (min/max: 30.13/84.65) |        |          | <b>94.73</b> (min/max: 88.10/100.00) |        |          |

*Bold values indicate the average accuracy after 10-fold testing.*

problem of dealing with large-scale time series and incorporating contextual knowledge to produce descriptive and predictive models with explanatory capability seems relatively unsolved at the time of this writing. In our preliminary experimentation, both AERA and Cisco's Kronos projects are demonstrating promising results. Incorporating AERA and Kronos into the hybrid architecture is expected to result in enhanced unsupervised learning and attention mechanisms directly from large-scale time series.

This evolved hybrid architecture (ML/DL/NARS/Kronos metamodel) is expected to promote cognitive synergy while preserving level of abstraction, symmetric and anti-symmetric properties of knowledge and using a bottom-up approach to refresh System-2 symbols from System-1 data integration (see Latapie et al., 2021 for details). Moreover, System-1 provides rapid responses to the outside world and activates System-2 in case of a surprise such as an emergency or other significant event that requires further analysis and potential action. System-2 uses conscious attention to request subsymbolic knowledge and sensory data from System-1, to be integrated into the levels of abstraction inspired from Korzybski's work. Korzybski's two major works (Korzybski, 1921, 1994) emphasize the importance of bottom-up knowledge. The corticothalamic and thalamocortical connections play different but complementary roles.

A balanced interplay between System-1 and System-2 is important. System-1's innate role is to ensure the many faceted health of the organism. System-2 is ideally used to help humans better contend with surprises, threats, complex situations, important goals, and achieve higher levels in Maslow's hierarchy of needs. From an AI systems perspective, contemporary Deep/Machine Learning methods (including Deep Learning) have it the other way around: Causal modeling and advanced reasoning are being solved in System 1, leveraging statistical models which can be seen as an inversion of proper thalamocortical integration.

## 5. CONCLUSIONS

While not conclusive, findings about natural intelligence from psychology, neuroscience, cognitive science, and animal cognition imply that both low-level perceptual knowledge and

higher-level more abstract knowledge may be neurosymbolic. The difference between high and low levels of abstraction may be that lower levels involve a greater amount of unconscious (automatic) processing and attention, while higher levels are introspectable to a greater extent (in humans, at least) and involve conscious (i.e. steerable) attention. The neurosymbolic metamodel and framework introduced in this article for artificial general intelligence is based on these findings, and the nature of the distinction between both systems will be subject to further research. One may ask whether artificial intelligence needs to mimic natural intelligence as a key performance indicator. The answer is yes and no. No, because natural intelligence, a result of billions of years of evolution, is full of imperfections and mistakes. Yes, because it is the best way known to help organisms survive for countless generations.

Both natural and artificial intelligences can exhibit astounding generalizability, performance, ability to learn, and other important adaptive behaviors when symbolic originating attention and sub-symbolic originating attention are properly handled. Allowing one system of attention to dominate, or inverting the natural order (e.g., reasoning in the subsymbolic space or projecting symbolic space stressors into the subsymbolic space) may lead to suboptimal results for engineered systems, individuals, and societies.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

HL and OK conceived of the presented idea and implemented the framework. HL designed the framework and the experiments. OK ran the tests and collected data. KT, PW, and PH contributed to the theoretical framework. All authors contributed to the writing of this manuscript and approved the final version.

## ACKNOWLEDGMENTS

The authors would like to thank Tony Lofthouse for his valuable comments.

## REFERENCES

- Alcaraz, F., Fresno, V., Marchand, A. R., Kremer, E. J., Coutureau, E., and Wolff, M. (2018). Thalamocortical and corticothalamic pathways differentially contribute to goal-directed behaviors in the rat. *eLife* 7:e32517. doi: 10.7554/eLife.32517
- Anticevic, A., Cole, M. W., Repovs, G., Murray, J. D., Brumbaugh, M. S., Savić, A. M. W. A., et al. (2014). Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. *Cereb. Cortex* 24, 3116–3130. doi: 10.1093/cercor/bht165
- Balleine, B. W., and Leung, R. W. M. B. K. (2015). Thalamocortical integration of instrumental learning and performance and their disintegration in addiction. *Brain Res.* 1628, 104–116. doi: 10.1016/j.brainres.2014.12.023
- Bengio, Y. (2019). "From system1 deep learning to system2 deep learning [conference presentation]," in *NeurIPS 2019 Posner Lecture* (Vancouver, BC).
- Bengio, Y., Lecun, Y., and Hinton, G. (2021). Deep learning for AI. *Commun. ACM* 64, 58–65. doi: 10.1145/3448250
- Bolkan, S. S., Stujenske, J. M., Parnaudeau, S., Spellman, T. J., Rauffenbart, C., Abbas, A. I., et al. (2017). Thalamic projections sustain prefrontal activity during working memory maintenance. *Nat. Neurosci.* 20, 987–996. doi: 10.1038/nn.4568
- Brannon, E. M. (2005). "What animals know about number," in *Handbook of Mathematical Cognition*, ed J. I. D. Campbell (New York, NY: Psychology Press), 85–108.
- Camp, E. (2009). "Language of baboon thought," in *The Philosophy of Animal Minds*, ed R. W. Lurz (Cambridge: Cambridge University), 108–127. doi: 10.1017/CBO9780511819001.007
- Clay, Z., and Zuberbühler, K. (2011). Bonobos extract meaning from call sequences. *PLoS ONE* 6:e18786. doi: 10.1371/journal.pone.0018786



- Diester, I., and Nieder, A. (2007). Semantic associations between signs and numerical categories in the prefrontal cortex. *PLoS Biol.* 5:e294. doi: 10.1371/journal.pbio.0050294
- Evans, J. S. B., and Elqayam, S. (2007). Dual-processing explains base-rate neglect, but which dual-process theory and how? *Behav. Brain Sci.* 30, 261–262. doi: 10.1017/S0140525X07001720
- Evans, J. S. B. T., and Stanovich, K. E. (2013). Dual-process theories of higher cognition: advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241. doi: 10.1177/1745691612460685
- Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207. doi: 10.1038/nature04675
- Gharghabi, S., Ding, Y., Yeh, C.-C. M., Kamgar, K., Ulanova, L., and Keogh, E. (2017). “Matrix profile viii: domain agnostic online semantic segmentation at superhuman performance levels,” in *2017 IEEE International Conference on Data Mining (ICDM)* (New Orleans, LA), 117–126. doi: 10.1109/ICDM.2017.21
- Gorban, A. N., and Tyukin, I. Y. (2018). Blessing of dimensionality: mathematical foundations of the statistical physics of data. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 440, 1204–1207. doi: 10.1098/rsta.2017.0237
- Goyal, A., and Bengio, Y. (2020). Inductive biases for deep learning of higher-level cognition. *arXiv preprint*. doi: 10.48550/arXiv.2011.15091
- Hammer, P., Lofthouse, T., Fenoglio, E., and Latapie, H. (2019). “A reasoning based model for anomaly detection in the smart city domain,” in *NARS Workshop in AGI-19* (Shenzhen), 1–10.
- Hart, D., and Goertzel, B. (2008). “Opencog: a software framework for integrative artificial general intelligence,” in *Proceedings of AGI2008* (Memphis, TN), eds P. Wang, B. Goertzel, and S. Franklin, 468–472.
- Hauser, M. D., Dehaene, S., Dehaene-Lambertz, G., and Patalano, A. L. (2007). Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (*Saguinus oedipus*). *Cognition* 86, B23–B32. doi: 10.1016/S0010-0277(02)00158-0
- Helgason, H. P., Thórisson, K. R., Garrett, D., and Nivel, E. (2013). Towards a general attention mechanism for embedded intelligent systems. 4, 1–7. doi: 10.5963/IJCSAI0401001
- Houwer, J. D. (2019). Moving beyond system 1 and system 2: conditioning, implicit evaluation, and habitual responding might be mediated by relational knowledge. *Exp. Psychol.* 66, 257–265. doi: 10.1027/1618-3169/a000450
- Hua, M., Chen, Y., Chen, M., Huang, K., Hsu, J., Bai, Y., et al. (2021). Network-specific corticothalamic dysconnection in attention-deficit hyperactivity disorder. *J. Dev. Behav. Pediatr.* 42, 122–127. doi: 10.1097/DBP.0000000000000875
- Hubbard, E. M., Diester, I., Cantlon, J. F., Ansar, D., van Opstal, F., and Troiani, V. (2008). The evolution of numerical cognition: from number neurons to linguistic quantifiers. *J. Neurosci.* 28, 11819–11824. doi: 10.1523/JNEUROSCI.3808-08.2008
- James, W. (1890). *The Principles of Psychology, Vol. 2*. New York, NY: Dover Publication. doi: 10.1037/10538-000
- Kahneman, D. (1973). *Attention and Effort*. New Jersey: Prentice-Hall.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.
- Keren, G. (2013). A tale of two systems: a scientific advance or a theoretical stone soup? cOmmentary on evans stanovic. *Perspect. Psychol. Sci.* 8, 257–262. doi: 10.1177/1745691613483474
- Kipf, T., Fetaya, E., Wang, K. C., Welling, M., and Zemel, R. (2018). Neural relational inference for interacting systems. *arXiv preprint*. doi: 10.48550/arXiv.1802.04687
- Koch, C., and Tsuchiya, N. (2006). Attention and consciousness: two distinct brain processes. *Trends Cogn. Sci.* 11, 16–22. doi: 10.1016/j.tics.2006.10.012
- Korzybski, A. (1921). *Manhood of Humanity, The Science and Art of Human Engineering*. New York, NY: E. P. Dutton and Company. doi: 10.2307/2972481
- Korzybski, A. (1994). *Science and Sanity: An Introduction to Non-Aristotelian Systems, 5th Edn*. New York, NY: Institute of General Semantics.
- Lake, B., and Baroni, M. (2018). “Generalization without systematicity: on the compositional skills of sequence-to-sequence recurrent networks,” in *Proceedings of International Conference on Machine Learning* (Stockholm), 2873–2882.
- Latapie, H., Liu, O. K. G., Kompella, R., Lawrence, A., Sun, Y., Srinivasa, J., et al. (2021). A metamodel and framework for artificial general intelligence from theory to practice. *J. Artif. Intell. Conscious.* 8, 205–227. doi: 10.1142/S2705078521500119
- Liu, D., Lamb, A., Kawaguchi, K., Goyal, A., Mozer, C. S. M. C., and Bengio, Y. (2021). Discrete-valued neural communication. *arXiv preprint*. doi: 10.48550/arXiv.2107.02367
- Llinas, R. R. (2002). “Thalamocortical assemblies: how ion channels, single neurons and large-scale networks organize sleep oscillations,” in *Thalamus and Related Systems*, eds A. Destexhe and T. J. Sejnowski (Oxford: Oxford University), 87–88. doi: 10.1016/S1472-9288(02)00034-1
- Marchetti, M. (2011). Against the view that consciousness and attention are fully dissociable. *Front. Psychol.* 3:36. doi: 10.3389/fpsyg.2012.00036
- Martin, C., Bhui, R., and Bossaerts, P. (2014). Chimpanzee choice rates in competitive games match equilibrium game theory predictions. *Sci. Rep.* 4, 51–81. doi: 10.1038/srep05182
- Miller, L. M., and D’Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J. Neurosci.* 25, 5884–5893. doi: 10.1523/JNEUROSCI.0896-05.2005
- Monteiro, S. M., and Norman, G. (2013). Diagnostic reasoning: where we’ve been, where we’re going. *Teach. Learn. Med.* 25, S26–S33. doi: 10.1080/10401334.2013.842911
- Newell, A., and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Commun. ACM* 19, 113–126. doi: 10.1145/360018.360022
- Nivel, E., Thórisson, K. R., Steunebrink, B., Dindo, H., Pezzulo, G., Rodriguez, M., et al. (2013). *Bounded Recursive Self-Improvement*. Tech Report RUTR-SCS13006, Reykjavik University; School of Computer Science. Reykjavik University, Iceland.
- Nivel, E., Thórisson, K. R., Steunebrink, B., and Schmidhuber, J. (2015). “Anytime bounded rationality,” in *Proceedings of 8th International Conference on Artificial General Intelligence (AGI-15)* (Berlin), 121–130. doi: 10.1007/978-3-319-21365-1\_13
- Noesselt, T., Riegerand, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., and Heinze, H. J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J. Neurosci.* 27, 11431–11441. doi: 10.1523/JNEUROSCI.2252-07.2007
- Papaioannou, A. G., Kalantzi, E., Papageorgiou, C. C., and Korombili, K. (2021). Complexity analysis of the brain activity in autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) due to cognitive loads/demands induced by aristotle’s type of syllogism/reasoning. A power spectral density and multiscale entropy (MSE) analysis. *Heliyon* 7:e07984. doi: 10.1016/j.heliyon.2021.e07984
- Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. (2019). Continual lifelong learning with neural networks: a review. *Neural Netw.* 113, 54–71. doi: 10.1016/j.neunet.2019.01.012
- Perry, J. C., Pakkenberg, B., and Vann, S. D. (2018). Striking reduction in neurons and glial cells in anterior thalamic nuclei of older patients with down’s syndrome. *BioRxiv* 2018:449678. doi: 10.1101/449678
- Posner, I. (2020). “Robots thinking fast and slow: on dual process theory and metacognition in embodied AI,” in *RSS 2020 Workshop RobRetro* (Corvallis, OR).
- Sampathkumar, V., Miller-Hansen, A., Sherman, S. M., and Kasthuri, N. (2021). Integration of signals from different cortical areas in higher order thalamic neurons. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2104137118. doi: 10.1073/pnas.2104137118
- Slovan, S. A. (1996). The empirical case for two systems of reasoning. *Psychol. Bull.* 119, 3–24. doi: 10.1037/0033-2909.119.1.3
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behav. Brain Sci.* 11, 1–43. doi: 10.1017/S0140525X00052432
- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., and Sedivy, J. C. (2013). Eye movements and spoken language comprehension: effects of visual context on syntactic ambiguity resolution. 45, 447–481. doi: 10.1016/S0010-0285(02)00503-0
- Steenkiste, S. V., Chang, M., Greff, K., and Schmidhuber, J. (2018). Relational neural expectation maximization: unsupervised discovery of objects and their interactions. *arXiv preprint*. doi: 10.48550/arXiv.1802.10353
- Strack, F., and Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Pers. Soc. Psychol. Rev.* 8, 220–247. doi: 10.1207/s15327957pspr0803\_1

- Sumner, P., Tsai, P. C., Yu, K., and Nachev, P. (2006). Attentional modulation of sensorimotor processes in the absence of perceptual awareness. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10520–10525. doi: 10.1073/pnas.0601974103
- Thórisson, K. R. (2009). *From Constructionist to Constructivist AI*. Tech Report FS-09-01. AAAI Fall Symposium Series: Biologically Inspired Cognitive Architectures, 175–183.
- Thórisson, K. R. (2020). “Seed-programmed autonomous general learning,” in *Proceedings of Machine Learning Research* (Cambridge, MA), 32–70.
- Thórisson, K. R., Bieger, J., Li, X., and Wang, P. (2019). “Cumulative learning,” in *Proceedings of International Conference on Artificial General Intelligence (AGI-19)* (Shenzhen), 198–209. doi: 10.1007/978-3-030-27005-6\_20
- Tyll, S., Budinger, E., and Noesselt, T. (2011). Thalamic influences on multisensory integration. *Commun. Integr. Biol.* 4, 145–171. doi: 10.4161/cib.15222
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Kaiser, A. N. G. L., et al. (2017). “Attention is all you need,” in *Proceedings of Advances in Neural Information Processing Systems* (Long Beach, CA), 5998–6008.
- Wang, P. (2005). Experience-grounded semantics: a theory for intelligent systems. *Cogn. Syst. Res.* 6, 282–302. doi: 10.1016/j.cogsys.2004.08.003
- Wang, P. (2006). *Rigid Flexibility: The Logic of Intelligence*. Dordrecht: Springer.
- Wang, P. (2013). *Non-Axiomatic Logic: A Model of Intelligent Reasoning*. Singapore: World Scientific. doi: 10.1142/8665
- Werner, S., and Noppeney, U. (2010). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb. Cortex* 20, 1829–1842. doi: 10.1093/cercor/bhp248
- Wolff, M., and Vann, S. D. (2019). The cognitive thalamus as a gateway to mental representations. *J. Neurosci.* 39, 3–14. doi: 10.1523/JNEUROSCI.0479-18.2018
- Xu, X., Hanganu-Opatz, I. L., and Bieler, M. (2020). Cross-talk of low-level sensory and high-level cognitive processing: development, mechanisms, and relevance for cross-modal abilities of the brain. *Front. Neurobot.* 14:7. doi: 10.3389/fnbot.2020.00007
- Yeh, C.-C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, A., et al. (2016). “Matrix profile i: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets,” in *2016 IEEE 16th International Conference on Data Mining (ICDM)* (Barcelona), 1317–1322. doi: 10.1109/ICDM.2016.0179

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Latapie, Kilic, Thórisson, Wang and Hammer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.