

NEURAL MODEL-APPLYING NETWORK (NEUMAN): A NEW BASIS FOR COMPUTATIONAL COGNITION

Frederick Roth

Professor (Retired), Information Sciences, Naval Postgraduate School, USA

ABSTRACT

NeuMAN represents a new model for computational cognition synthesizing important results across AI, psychology, and neuroscience. NeuMAN is based on three important ideas: (1) neural mechanisms perform all requirements for intelligence without symbolic reasoning on finite sets, thus avoiding exponential matching algorithms; (2) the network reinforces hierarchical abstraction and composition for sensing and acting; and (3) the network uses learned sequences within contextual frames to make predictions, minimize reactions to expected events, and increase responsiveness to high-value information. These systems exhibit both automatic and deliberate processes. NeuMAN accords with a wide variety of findings in neural and cognitive science and will supersede symbolic reasoning as a foundation for AI and as a model of human intelligence. It will likely become the principal mechanism for engineering intelligent systems.

KEYWORDS

Neural Nets, Cognitive Models, Symbolic Reasoning, Holographic Memory, Automaticity.

1. BACKGROUND AND MOTIVATION

Since the 1950s researchers have investigated what intelligence means and how to mechanize it. From the outset three very general capabilities dominated consideration: learning, automaticity, and reasoning. Learning encompasses memorization and recall, abstraction, composition and hierarchical modeling, among other capabilities. Automaticity describes the ability of humans to perform well learned behaviors smoothly and unconsciously, as when adults read learned words and phrases without any awareness of effort or component subtasks. Reasoning encompasses problem solving, planning, and logical inference.

Cognitive psychologists, neuroscientists, and AI researchers, over the decades, increasingly focused on the key role of models in supporting important functions such as understanding, prediction, and planning, among others.[1-5] Speech and language models underlie the processing of natural language, and humans use a wide variety of additional models to assess their situations and execute plans appropriate to those situations.

Progress in employing symbolic reasoning for these tasks has stagnated, for several reasons. First, algorithms to learn symbolic rules and to choose which rules to fire require combinatorial methods characteristic of NP-complete problems.[5, 6] Symbolic reasoning systems employing von Neumann computers cannot attain automaticity, because matching situations to conditions incurs delay while binding symbols to corresponding objects.

Recent neural net (NN) studies have shown that, with broad dendritic trees, sufficient layers, suitable sensor manifolds, convolutional and recurrent layers, modern capable NNs can eventually achieve most of the long-standing objectives for intelligent systems. The goal of this paper is to describe how, with a few extensions, an enhanced NN design can achieve all the objectives for intelligence. We refer to the proposed construct as the Neural Model-Appling Network (NeuMAN). It aims to supersede von Neumann computer software architectures as a basis for future progress in intelligent systems and cognitive psychology.

2. MEMORY, ABSTRACTION, PATTERN LEARNING, COMPOSITION AND HIERARCHICAL MODELING

Intelligent agents must remember what they have experienced, and from those experiences they need to infer generally valid patterns and rules. We denote by $S(t)$ the engram experienced by the agent at time t . The engram comprises the full state of the agent's computing basis and, in the case of humans, this typically means the activity of all neural components. For simplicity, we assume every neural cell corresponds to one feature that is either ON or OFF at time t , and the engram consists of the set of ON cells. The agent learns classes by abstracting the common elements of examples of each class. As a side-effect of classification learning, the agent exhibits generalization across stimuli that have most of the defining characteristics of a learned class. Abstraction and generalization reflect two features of the same learning process.

An engram $S(t)$ at time t is followed by the engram $S(t+1)$ at time $t+1$. In any NN, active cells at time t cause some cells to fire at time $t+1$, and this provides a basis for abstracting patterns across sequences of successive engrams. Abstractions across adjacent states $\langle S(t), S(t+1) \rangle$ constitute inferred rules of the form $S^*(t) \rightarrow S^*(t+1)$, meaning that the common features of consequent states $S(t+1)$ will follow antecedent states manifesting all features in $S^*(t)$.

More generally, the agent memorizes longer sequences of successive states, such as $\langle S(t), S(t+1), \dots, S(t+k) \rangle$ for some small value of k . These memorized sequences provide the basis for learning abstracted patterns, $S^*(t) \rightarrow S^*(t+1) \dots \rightarrow S^*(t+k)$, and we denote these as $k+1$ tuples $\langle S^*(t), S^*(t+1), \dots, S^*(t+k) \rangle$. Each such pattern specifies a set of features present in each successive engram. Situations that match the abstraction $S^*(t)$ will likely match the abstraction $S^*(t+1)$ at the next instant, and so on, until the k -th successive event, learned to match $S^*(t+k)$.

Learned patterns of the sort just described could be denoted as P_j , for some j . When a learned pattern P becomes activated at time t , P becomes part of the engram $S(t)$. Thereafter, newly learned patterns can incorporate P in their description. This is important for two reasons. First, patterns can include other patterns, as pattern learning creates hierarchical compositions of sub-patterns. Pattern learning and composition produce higher-level patterns. The expectation that some pattern $Q(t+1)$ follows $P(t)$ in a learned pattern $\langle \dots, P(t), Q(t+1), \dots \rangle$ means that the components of $Q(t+1)$ will occur starting at $t+1$. Patterns of patterns correspond to hierarchical models of expected states.

Patterns can also describe action sequences when they designate effector cells. When components of these patterns specify agent actions, these patterns constitute hierarchical plans whereby the agent successively enacts component actions or subplans, recursively.

Composition of abstracted patterns produces models of sensed states and actions at successively higher levels. High-level models efficiently encode situations and plans that the agent reasons about. These encodings may be specific for describing locations, one which is agent-centric for "where" actions and one which is object-centric for "what" actions. [5].

Previous research in both symbolic reasoning and NNs has shown capabilities for memorization, abstraction, pattern learning, composition, and hierarchical modeling. The advantages of employing symbolic reasoning to accomplish such learning include: (1) we ordinarily produce a small number of learned patterns; (2) the learned patterns comprise a small number of user-defined features and previously inferred sub-patterns; and (3) we can easily test and validate algorithms have been developed to operate with such data and produce symbolic descriptions of classes and patterns.

NNs, on the other hand, have been shown capable of memorizing engrams, learning patterns, and abstracting hierarchical models. [8-10]. Patterns learned by NNs, however, resist efforts to characterize them succinctly as we can with symbolic abstractions. [11-13] NNs operate upon an underlying representation system that differs in key ways from symbolic representations: (1) no symbols or binding functions exist; (2) all engram features correspond to states of particular neurons, such as ON or OFF or other similar discretizations of neuron firing activity; and (3) learning reinforces every possible abstraction in memory, and compares firing strengths of alternative patterns to determine which features and events trigger subsequent activity. In short, NNs represent situations as conjunctions of features based on underlying dimensions such as space, time, frequency and intensity, whereas symbolic reasoning uses variables as placeholders for objects that, when found, satisfy the relations defining a class or matching a pattern. NNs employ manifolds of sensing cells with overlapping receptive fields so reinforcement strengthens all feature subsets needed to characterize a learned concept. [14-16] Convolutional NNs (CNNs) apply filters across these manifolds to extract features for higher-level analysis. Recurrent NNs (RNNs) essentially replicate sensory manifolds at subsequent levels to support learning of temporal patterns. Thus, NNs can learn sequential patterns and patterns invariant across common transformations.

OpenAI's Brown, *et al.* [17] demonstrated that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art fine tuning approaches. They trained GPT-3, an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model had at the time and tested its performance in the few-shot setting. For all tasks, GPT-3 was applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified via text interaction with the model. GPT-3 achieved strong performance on many NLP datasets, including translation, question-answering, and cloze tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic. GPT3 could generate samples of news articles which human evaluators have difficulty distinguishing from articles written by humans. Large language models such as GPT3 and its many successors can be mapped onto many types of parameter objects. For example, Dalle-2 [17] and Midjourney [18] can produce detailed synthetic images in response to text prompts. An artist named Jason Allen recently utilized Midjourney to win a controversial First Place in a Colorado State Fair Arts Competition in the "digital arts/digitally manipulated photography" category.[19]. DeepMind's AlphaFold predicted the 3D structure of virtually all 200 million known proteins, expanding the number of known 3D protein structures from 190 thousand. [20] This is a huge scientific and medical achievement that has already resulted in DeepMind researchers winning one of the \$3M Breakthrough prizes.[21]

Researchers in the fields of cognitive psychology and AI have shown that, across a wide variety of domains, both symbolic reasoners and NNs can memorize and abstract features common to training examples, learn patterns, and compose patterns into hierarchical models. Each approach has also revealed shortcomings. Symbolic reasoning compels us to express knowledge in terms of named relations applied to variables referring to potential matching objects present in the engram, and this entails an NP-complete matching algorithm that cannot avoid temporal delays in binding

objects to variables.[6] Moreover, learners must reinforce an exponential number of candidate patterns and abstractions inferable from training data, because every possible subset of features justifies another inference.[22, 23] In contrast with von Neumann learners, NNs with deep learning strengthen every appropriate synapse involved in each reinforced pattern, and this results in the NNs strengthening all possible inferences in parallel. NNs typically build in low-level features across a manifold aligned with the dimension associated with each kind of feature, such as position, frequency, intensity, orientation, ON or OFF, and so forth. Convolutional NNs (CNN) employ a manifold across adjacent manifolds to recognize instances of the same pattern independent of its position in the environment. Research has shown these CNNs capable of learning all types of familiar patterns, such as characters, sounds, faces, and natural language. Rather than encoding these patterns in concise abstract terms, the NNs learn precisely how to discriminate instances of a pattern through as many successive layers as required. NN have virtually unlimited capacity and discriminative capability.[24,13] For example, Deep Mind's Gato is described as a "Generalist Agent". Gato uses multi-modal, multi-task, multi-embodiment generalist policy. One network with the same weights can play Atari, caption images, chat, stack blocks with a real robot arm and more, deciding based on context whether to output text, joint torques, button presses, or other tokens. It performs over 450 out of 604 tasks at over a 50% expert score threshold.[25]

For these reasons, NeuMAN forgoes symbolic reasoning and utilizes a NN foundation as a basis for intelligence with automaticity. In addition, agents can perform procedures, moving from one step to the next automatically. When the agent performs such procedures, it operates upon mental models based in NeuMAN and represented as engram feature sets.

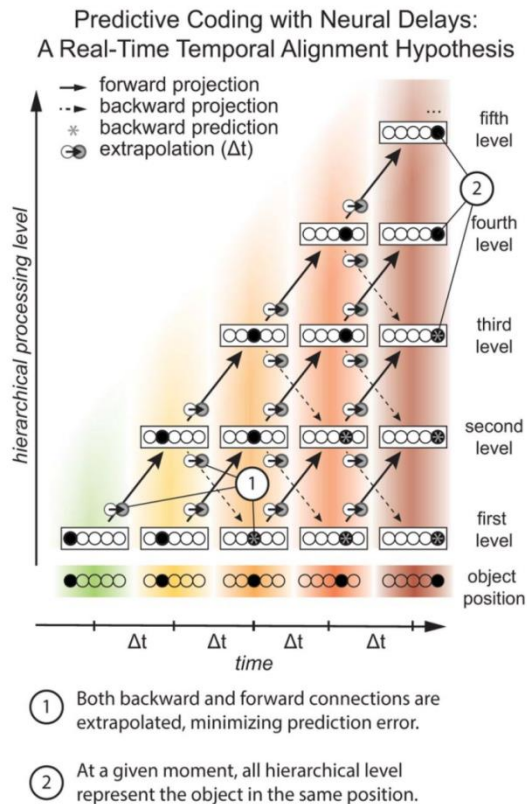


Figure 1. Multilevel networks with neural delays learn patterns and make forward and backward predictions to refine and strengthen correct hypotheses. [27]

3. PREDICTION, EXPECTATION, AND INFORMATION VALUE

A learned sequential pattern such as $S^*(t) \rightarrow S^*(t+1)$ enables the agent to predict that features in $S^*(t+1)$ should occur in the $S(t+1)$ engram following the occurrence of features in the $S(t)$ engram matching the abstraction $S^*(t)$. For example, one might expect to perceive the word “House” immediately following “White” based on the frequently occurring word sequence \langle “White”, “House” \rangle . Because the agent learns longer sequences and higher-level composable patterns, an American speaker of English ultimately learns other patterns providing additional context for accurate predictions, such as predicting the White House as a location where the US President works and resides. In a context of reporting about politics, this makes “House” a very strong prediction from the previous word “White.” In a context about cars for sale, however, “white” likely refers to an automobile paint color. Predictive strength, termed diagnosticity [23, 26], has been cited by many researchers as a key determinant in what agents notice and respond to. Because all information processing systems face constraints on resources that limit computing speed, long processing delays impair ideal performance. NeuMAN responds first to the strongest competing signals, allowing weaker alternatives to languish.

Many papers in neuroscience and natural language studies describe detailed mechanisms for pattern learning and prediction. We cite two examples here to illustrate the underlying mechanisms and resulting behavior. Hogendoorn and Burkitt [27] describe how multilayered neural networks essentially represent temporal patterns across successive layers acting as a shift buffer. Figure 1 above from their paper describes this process.

Another paper [28] illustrates pattern learning and prediction in natural language learned by a NN. In this paper, natural language categories and parts of speech emerged from unsupervised training on word sequences, with the network learning how to predict prior and subsequent words. Figure 2 illustrates their findings that this simple task generates a vast amount of syntactic and semantic knowledge.

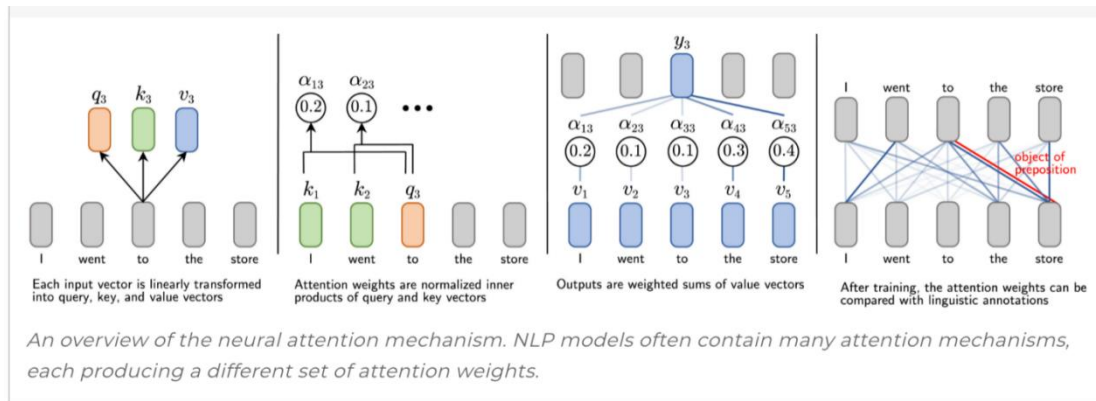


Figure 2. Words and word sequences learned as patterns enable accurate predictions of prior and subsequent words as well as identifying syntactic and semantic categories. [28]

Predictive strength increases with the diagnosticity of the predictor, the strength of the predictor itself, and the learned importance of the predicted feature. The diagnosticity of a predictor p for a predicted consequent q is related to the conditional probability of q given p , and this is proportional to the likelihood of q following p . When we say a feature p is present, we mean its firing strength exceeds some established threshold. We can make this binary, so that the strength of any feature is 1 or 0, or we can use more refined scales. Similarly, we can model the learned importance of a feature as 1 or 0, although in keeping with deep learning results, we believe finer

granularity will produce better results. Given a strong predictor of a strong feature q , we strongly expect q will fire.

Predictions have been shown to play many useful roles in various systems. Accurate predictions allow the agent to anticipate subsequent events and, ideally, reduce errors and energy by exploiting the predictions. Predictions are specific to a reference frame, whether it is at the level of a finger's touch, a sound, a plan, or the location of a previously stored object in a room. Regardless of the specific implementation, a strong prediction should reduce or obviate efforts subsequently expended to sense and respond to anticipated features. On the other hand, when sensed events contradict predictions, these surprises should generate more activity than they would otherwise. Elsewhere, we have shown how surprises like this constitute valued information at the right time (VIRT) [29]. Systems incorporating VIRT principles to suppress processing of expected results and quickly focus on surprising information reduce information processing loads by orders of magnitude.[30, 31] The value of information, from this perspective, measures the importance of a quick response to the reported event or, equivalently, the cost of a delayed response. Events that don't require any response have minimal information value. Correctly predicting events enables the agent to confirm its plans and situation analyses with minimal effort. This, in turn, means the agent has more resources available for other activities.

Neuroscientists have long understood the value of predictive coding. As de-Wit,*et al.*[32] state: "Predictive coding posits that the brain actively predicts upcoming sensory input rather than passively registering it. Predictive coding is efficient in the sense that the brain does not need to maintain multiple versions of the same information at different levels of the processing hierarchy." In the predictive coding model of [33], predictions generated at higher levels are used to "explain away" compatible and redundant lower-level representations. This explaining away reduces activity in early areas through feedback from higher-level areas. Other studies using fMRI have confirmed this finding.

Huang and Rao [34] summarize the neuroscience findings in this way:

Predictive coding is a unifying framework for understanding redundancy reduction and efficient coding in the nervous system. By transmitting only the unpredicted portions of an incoming sensory signal, predictive coding allows the nervous system to reduce redundancy and make full use of the limited dynamic range of neurons. Starting with the hypothesis of efficient coding as a design principle in the sensory system, predictive coding provides a functional explanation for a range of neural responses and many aspects of brain organization. The lateral and temporal antagonism in receptive fields in the retina and lateral geniculate nucleus occur naturally as a consequence of predictive coding of natural images. In the higher visual system, predictive coding provides an explanation for oriented receptive fields and contextual effects as well as the hierarchical reciprocally connected organization of the cortex. Predictive coding has also been found to be consistent with a variety of neurophysiological and psychophysical data obtained from different areas of the brain.

NeuMAN operationalizes these ideas by using predictions to generate expectations. The effects of expecting an event e depend on subsequent events. When the expected e occurs, e fires at a strength less than what would occur had the agent not predicted it. Thus, expectations result in inhibition of the associated cells. In addition, ongoing reinforcement strengthens the activated cells and synapses that produce confirmed expectations. Thus, the agent reinforces antecedent cells and synapses producing valid expectations and reduces responsiveness of successor cells recognizing expected features. Basically, the network models on-going predictable activity while minimizing effort and attention. Various papers in the literature show that neural nets use expectations in the way described here. As an excellent example [35] states:

In particular, expectation has been defined as the (implicit or explicit) knowledge of the probability of occurrence of a stimulus, independent of its task relevance. Attention, by contrast, refers to the relevance of a stimulus for an upcoming task, independent of its probability. Defined in this way, expectation and attention have distinct effects on stimulus-evoked visual cortex activity: attended stimuli elicit stronger responses than unattended stimuli, whereas expected stimuli elicit weaker responses than unexpected stimuli. In particular, expectation has been defined as the (implicit or explicit) knowledge of the probability of occurrence of a stimulus, independent of its task relevance. Attention, by contrast, refers to the relevance of a stimulus for an upcoming task, independent of its probability.

As another example consistent with those findings [36] shows how expectations inhibit responses so the agent can ignore predictable goal-irrelevant or distracting information.

NeuMAN learns to predict sequential features, mostly ignore insignificant events, and favor activities that respond to high information value.

4. REINFORCEMENT LEARNING FOR SENSING AND ACTING

Consistent with extensive research in psychology and AI, NeuMAN relies on reinforcement of behaviors producing desirable results to strengthen synapses and cells contributing to the positive outcomes. NeuMAN incorporates these methods and applies them ubiquitously. This makes reinforcement the primary determinant of what an agent learns given its endowment of built-in features and network wiring [37]. Given an initial endowment of sensed features and executable behaviors, the agent responds to successive situations by recognizing features and activating the strongest associated cells and their corresponding components. Knowledge of two sorts accumulates in response to reinforcement. First, the agent learns patterns that address sensing, and these patterns become the hierarchical models the agent employs to assess its situation. Second, the agent learns patterns that enable it to activate appropriate responses and to expect anticipated sensations that it can mostly ignore.

In supervised training of deep NNs, explicit reinforcing signals strengthen active pathways producing the desired response. Synapse change rules strengthen connections between antecedent and consequent cells and assemblies. In animal models primary reinforcers such as food and sex respond to hardwired drives, such as hunger and lust. Secondary reinforcers get their potency through association with primary reinforcers, usually through operant conditioning where the agent has learned the diagnostic value of the secondary reinforcer as a predictor of an eventual reinforced outcome. In short, outcomes the agent experiences as positive strengthen the pathways that predictably attain those outcomes.

We know from a wide variety of studies that both people and NNs acquire models of sensed information at various levels of abstraction, as when spoken sounds are parsed at phonological, syllabic, lexical, syntactic, semantic and pragmatic levels. Figure 3 illustrates a sensory hierarchy more generally, ranging from base models of physical inputs, to situation models and possibilities used in situation assessment, planning, and reasoning, and culminating in executive functions such as quickly recognizing a critical condition requiring an immediate change in behavior. The figure shows that every level of sensing has a corresponding level of acting, whereby the agent determines how to behave. At the lowest level, the agent signals its effectors to operate muscles. Natural entities incorporate drives to seek food, water, shelter, relationships, and so forth. These drives combine with the sensed situation to prompt the agent to focus attention on high-value goals and to choose plans most appropriate to achieving those goals. Plans ultimately incorporate sequences of actions that produce behavior.

Figure 3 aims to convey several ideas without claiming precision. First, intelligent agents learn and apply hierarchical models at various levels. Higher-level models emerge as a consequence of deep learning in multilevel NNs. Higher-level sensing models have more value for the agent, because they reduce its information processing requirements. Expectations from these models inhibit lower-level responses. In addition, intelligent agents can focus attention on important aspects of modeled situations to assure they seize opportunities and avoid threats quickly. When an agent recognizes that its current situation warrants a known response, it can immediately trigger that plan without additional cogitation. With sufficient reinforcement, these responses strengthen and ultimately become automatic. Many high-value systems have been developed that couple situation assessment with planning and control [38-42]. Researchers have made progress on identifying a variety of multi-level sensing and acting models. Models have been learned through deep learning for a wide variety of environments, including speech, text, images, objects, and others. Psychologists have used multilevel models to describe a wide range of human perceptual tasks, especially situation analysis, planning, and control. [43-48] Figure 3 presents a rough overview of that literature, emphasizing that intelligent agents seek positive outcomes in response to reinforcements and internal drives, while modeling both the external world and the agent's own internal state.

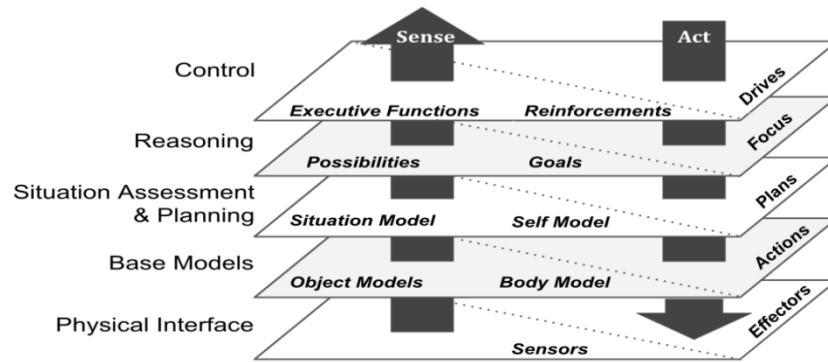


Figure 3. An intelligent agent must sense and act, and it uses diverse sorts of models to assess its situation and choose appropriate actions.

Higher-level models emerge automatically and inexorably from deep learning in multi-level NNs. Agents with sufficient endowments of sensors will naturally acquire mental models, including models of self. Models of one's own state will produce awareness, including awareness of one's own mental models for situations, possibilities, goals and choices.

A recent paper by Eppe *et al.*[49] surveys the cognitive psychology literature to propose that integrated hierarchical reinforcement learning will produce intelligent problem-solving in networks of sufficient capability. They highlight, specifically, the importance of compositional abstraction and predictive processing, as we have. Their results “suggest that all identified cognitive mechanisms have been implemented individually in isolated computational architectures, raising the question of why there exists no single unifying architecture that integrates them.” They argue that biological mechanisms, including forward and inverse models, intrinsic motivation, compositional abstractions, and mental simulations underlie learned hierarchical models for sensing and acting, as described here and pictured in Figure 3.

NNs of sufficient capability will learn and adopt models of their environments, their internal states, and their potential actions. These machines will recognize pertinent aspects of their situations, and those will necessarily incorporate models of all reinforced patterns, including internal states. When the agent deliberates about its own state, considers alternative courses of

action, and chooses one of these, the agent exhibits self-awareness. Awareness, in our minds, differs from sentience, because it does not require feeling or emotions, and purely textual learners lack actual sensations of those sorts. As in the Turing Test, however, a purely textual conversationalist can simulate what feelings humans would articulate. In this way, learning agents can generate behaviors from their inferred models of how people describe feelings and emotions. These simulations can fool observers into attributing sentience to the agent when the agents have persuasively modeled conversations about topics with emotional content.

Sufficiently capable NNs will learn models of whatever we train them on, and NeuMAN will couple models of sensing with models of acting to produce intelligent behaviors from the highest-level plans appropriate. Increasingly, NeuMAN machines will exhibit awareness, competence, and information processing efficiency. Oaksford and Chater [50] posit that the human brain maintains a probabilistic cognitive model of the world, and NeuMAN accords with their assessment.

5. HOLOGRAPHIC MEMORY (HM) AND INTERFERENCE MATCHING (IM)

Researchers have postulated that human brains function, in part, as holographic memory stores. The original holographic memories recorded the interference pattern between an image to be learned with a reference beam. After this recording, using the same reference beam to illuminate the storage medium reconstructs and displays the learned image. Other researchers postulated that the brain could create and recover memories in similar ways. Pribram [51] hypothesized that the brain used electrical patterns to provide holographic storage and retrieval. Cavanagh [52] showed that synaptic change rules could enable the brain to create and retrieve memories from the interference of dynamic situation engrams with periodic reference waves.

More recently, NN researchers have shown that deep learning synapse change rules can store and recall images essentially perfectly [13, 24]. These results show that NNs with wide trees, multiple levels, and appropriate synapse updating can effectively record and recall every reinforced experience. NeuMAN systems will be implemented in a variety of memory implementations that will grow in capability over time providing a virtually limitless memory with holographic capabilities.

Analysis of what NNs learn from large training sets, such as those for game playing or language learning, indicates that NNs learn and store specific examples and abstracted patterns, as well as learning categories of substitutable items such as all the elements of each part of speech. Thus, the English language learner implicitly learns determiner, for example, by finding that words a, an, the, one, and so forth can all precede the implicitly learned categories of adjective and noun and can also succeed occurrences of words belonging to category of transitive verbs. Each element of a category has similar diagnosticity for its preceding and succeeding category elements in learned sequential patterns.

Higher-order patterns comprise sequences of learned categories and patterns, recursively. Language learning occurs from a rich corpus of examples, the learner acquires patterns that explain and predict what was heard and what should be said. Learners master linguistic categories and patterns as well as irregularities and special cases, all through synapse change rules responding to reinforcement.

We have thus far over-simplified language learning by treating language merely as text. But spoken language has many other features, including pitch and prosody, and these convey additional meaning, such as the emotional state of the speaker. Learning to model how the speaker feels enables us to understand better what speakers mean when they say such things as

“No way” or “Heads up!” or “PLEASE!”. All language users rely on context, such as the speaker’s current situation and the sequence of words spoken. NeuMAN machines will learn across all conditions expressed in the corresponding engrams.

Sufficiently capable NNs will learn models of whatever we train them on, and NeuMAN will couple models of sensing with models of acting to produce intelligent behaviors from the highest-level plans appropriate. Increasingly, NeuMAN machines will exhibit awareness as well as competence. Strongly reinforced responses to expected situations will become automatic, leaving additional resources available for exceptional conditions.

6. PRINCIPAL CAPABILITIES OF HOLOGRAPHIC MEMORY (HM)

The capabilities of HM provide the foundation for intelligence. There are six primary capabilities:

1. The agent strengthens memory for every engram and all constituent assemblies it experiences in response to reinforcement.
2. The agent activates every learned pattern, or cell assembly, matched by the current engram, and the augmented engram explicitly includes the pattern as a constituent feature.
3. The agent reintegrates memorized assemblies when a diagnostic portion of its elements become active.
4. The agent memorizes every one-step sequential pattern $\langle S(t), S(t+1) \rangle$ and every experienced pattern $P = \langle P1, P2, \dots, P|P \rangle$ in response to reinforcement.
5. The agent predicts (expects) succeeding elements of patterns whose initial elements are active and postdicts (expects) preceding elements of patterns whose later elements are active. [53, 54]
6. The agent responds to the most strongly activated assembly among all mutually exclusive competitors.

Thus, HM provides for storage and recall, as well as learning of reinforced constituent assemblies and sequential patterns.[23, 56] Previous papers made many similar assumptions, but avoided the assumption of effectively total memory for reinforced experiences. Earlier, symbolic reasoning learning algorithms arose from the same assumptions, but these always relied upon algorithms to winnow the exponentially large populations of potentially learned elements. Genetic algorithms [56] and schematic classifiers [22] best exemplify approaches to maximize learning of exponential numbers of candidates with only a finite set of learned classifiers.

In contrast to the resource-constrained learning algorithms for von Neumann computers, NNs offer a mechanism for near-total, near-perfect memory of training examples. Synapse change rule adjustments have proved adequate. Competition among firing cells determines which assemblies and patterns prevail. NeuMAN memories activate all matching assemblies and patterns, but the strongest assemblies determine the agent’s response. These systems can be implemented with suboptimal off-the-shelf hardware and software. Next-generation parallel processing systems such as 3D protonic programmable resistors will increase processing speeds by four orders of magnitude and reduce power requirements dramatically [57].DeBole and team [58] may further accelerate the speed of pattern matching with a brain-inspired neuromorphic computing architecture that demonstrated a massively parallel neural inference engine consisting of a million spiking neurons and 256 million synapses. Neuromorphic computing architectures have the potential to revolutionize the speed of brain-inspired processor chips and decrease power requirements. Once NeuMAN class architectures become the norm, there will be a competitive ecosystem of hardware and software implementations available.

7. PRINCIPAL CAPABILITIES OF INTERFERENCE MATCHING (IM)

Hayes-Roth [59] introduced the term *interference matching* to describe the process for comparing situation descriptions to abstract common patterns. IM identifies the commonalities and differences between two or more descriptions. IM provides the essential capability for learning by abstraction, where all members of some class must share identical partial descriptions. The partial description common to all examples defines the set's abstraction. Multiple alternative abstractions always exist. The best model of every class usually comprises the simplest abstraction with the greatest diagnosticity. In the realm of symbolic reasoning, our algorithms maintain a population of candidate abstractions for every class and do not explicitly enumerate all subsets of those abstractions. More details on the use of IM on symbolic representations appear elsewhere.[6, 22, 53, 59],

Before the advent of modern capable NNs, the AI field focused on reasoning with symbolic descriptions using von Neumann computers. Those efforts proved valid and powerful for decades, but advances with NNs mark a turning point. NNs have become the primary mechanism for learning and memory. NeuMAN learns and applies models without variables and thus obviates sequential matching. In this context, IM operates ubiquitously and rapidly, exploiting the symbol-free neural manifolds and layers to represent situations on finite, explicit descriptions composed of active cells and assemblies.

IM puts two matched inputs into a correspondence, maximizing the common components and identifying the ways each input differs from that common abstraction. With NeuMAN as a base mechanism, IM performs instantaneously. Two activated memories M and N will activate most strongly the components common to both. Because M and N comprise sets of active cells, the elements common to both constitute the set intersection, $M \cap N$, and the differences unique to M and N, correspond to $M - M \cap N$ and $N - M \cap N$, respectively. When we say that the HM learns from all reinforced experiences, we mean that whenever M or N is reinforced, their common abstraction $M \cap N$ is also reinforced. Reinforcement, as in deep learning, uses feedback to strengthen the activated pathways that produced the reinforced behavior. The frequency of occurrence of any abstraction such as $M \cap N$ will always exceed the frequency of M or N, so when an abstraction proves diagnostic for a reinforced behavior, that abstraction strengthens more than any of the differences.

NeuMAN eliminates the need for IM operating on symbolic descriptions with von Neumann algorithms. IM on symbolic descriptions with variables is an NP-complete problem. In contrast, IM on neural engrams is an instantaneous process capable of producing fast and automatic responses. NeuMAN eliminates symbolic matching and slow computations for learning, for matching situations to learned conditions, and for invoking associated learned responses.

7.1. Examples of IM Use

Earlier papers showed that IM underlay the learning of patterns and rules. The advent of capable NNs mostly obviate symbolic approaches to IM, however. On the other hand, once armed with an HM capable of IM, we can immediately see that IM powers cognition in various ways. The following sections illustrate how NeuMAN would enable IM for everyday thinking and problem solving. Once we accept that HM engrams store and recall all reinforced experiences and sequential patterns, we gain a new perspective on cognition. IM processes directly enable a wide variety of intelligent behaviors. NeuMAN may provide the current best model of human cognition, because it accords with a vast amount of data and makes many new, plausible, vulnerable predictions.

Using IM, applying it to HM, enables the agent to perform many information processing functions quickly. These same functions simply cannot be performed quickly using symbolic representations. These rapid solutions result from the extensional, variable-free models of NeuMAN engrams. This capability of IM in HM seems analogous to the analog approach suggested for quickly sorting a set of alternatives to find the maximal element: First, adopt one rigid strand of spaghetti for each item to be sorted, with its length indicating its sort value. Then, grasp the entire bunch of noodles vertically and tap all their bases simultaneously on a table top. This instantly sorts the entire set, and the maximal element emerges immediately as the one standing taller than the rest. In an analogous way, IM in NeuMAN can identify all matching experiences and patterns in HM and respond to the strongest activated ones in parallel.

The paragraphs below illuminate some of the important cognitive capabilities IM in NeuMAN enables.

7.2. Seeking Examples and Counter Examples

When we have a conjecture or need to support a claim, we must find examples to illustrate our position. If we conjecture that automobiles have tires, we use the learned label “automobiles” to activate images and models of automobiles and verify that all of those match the model associated with the label “tires.” That approach produces the typical fast and affirmative response. However, people trained in statistics and logic, have learned that such an approach can lead to errors [60]. The trained reasoner knows that the real task is to determine if any example of an automobile lacks tires. Because NeuMAN learns from active features, we would not ordinarily notice or record “no tires” as a feature. We cannot use an absent feature for associative recall. The absence of a feature, such as “tires,” means it will not occur in engrams, will not occur in learned memories, and won’t enable recall. Knowing this, the skilled thinker will try to retrieve counterexamples using noticed features, as described below.

Counterexamples to general claims enable the agent to avoid over generalizing and to hone categories and inferences precisely. Continuing with the prior illustration, an agent looking for cars without tires needs to search memory with positive features rather than absent features. This can be done by the agent substituting various conjectured features for the missing tires, and then searching memory for examples. In this case, the agent might conjecture that the car’s metal wheels operate directly on the asphalt surface. Going further, the agent might conjecture that the car’s wheels operate directly on a different substrate. This may retrieve memories of cars and trucks operating with their wheels directly on metal train rails.

An ability to seek and find counterexamples greatly increases the speed and quality of learned patterns, especially conjectures about classifications and inferences. When memory retrieval fails to find examples of the sought pattern, the agent can employ a deliberate process to conjecture that such a counterexample occurs and seek ways to construct it. In such a way, the agent can explore new models and direct efforts to expand and refine its knowledge.[61-63]

7.3. Differentiating Two Sets

Often agents need to determine how and why two sets differ. For example, we might want to determine how candidate products differ. We routinely sample experiences with the two products, then use IM to find what’s common to each set and determine how the two examples with a common abstraction differ. We do this when we try on various products, such as clothes, eyeglasses, audio speakers, TV sets and so forth. Much of our knowledge rests on correctly classifying examples and avoiding misclassifications. Deep NNs learn to produce reinforced

responses, each appropriately tied to sensed features of alternative classes.

Beyond this automatic classification-based learning, an agent can use IM deliberately to match the common abstractions of two sets against each other. IM identifies the ways the two sets are the same and different. When performed deliberately, identified differences provide a conjectured basis for predicting why and how situations in the two cases evolve differently. All machine learning algorithms exploit such differences. IM in HM makes it quick and easy to differentiate two sets of memories.

7.4. Conjecturing Concepts and Categories

Research in AI and psychology has focused on concept learning and classification for decades. IM generates all abstractions of a training set, and those become candidates for identifying additional members of the concept. In this way, we learn to recognize symbols, images, words, and so forth. As discussed earlier, when we learn sequential patterns, these often include placeholders for categories that comprise substitutable members of that category. In language, for example, these categories include the familiar parts of speech, among others [64-65]. A partially matched pattern will predict and expect other elements of the pattern, and when one of those elements corresponds to such a category, the agent expects that category and all of its members. As an example from natural language learning, Determiner predicts Adjective and Noun Phrase and postdicts Transitive Verb, while Adjective postdicts Determiner, Adjective and Adverb and predicts Adjective and Noun Phrase, etc.

Applying IM to a set of examples generates a description of what's common and what's different across the examples. IM applied to sequences of natural language will find high agreement across syntactically similar word sequences, and these will identify categories and their members. Specifically, the set of alternatives that can appear in one place within a reinforced pattern constitute an implicitly defined category. Subsequently, that category becomes a feature present in associated engrams. The category is predicted by other elements of that pattern. Activation of any member of the category activates the corresponding feature. As a category provides a higher-level model of situations, NeuMAN reduces effort and computation at the lower-level features subtended by the category feature itself.

7.5. Refining a Concept

Intelligent agents act implicitly or explicitly on knowledge, especially which classes warrant which responses. So class concept definitions constitute the heart of one's knowledge. For this reason, the literature of psychology and machine learning has focused extensively on classification learning. We can focus on the everyday need to adjust one's concepts in light of an error. Typically, the error arises because the agent has chosen an inappropriate response to a situation matching one of its antecedent concepts associated with the chosen incorrect response. Methods for adjusting concepts in these cases are described in several places, including [64-68]. The need to refine a concept arises when a matching situation produces an incorrect response. The agent needs to refine the antecedent condition to block the current situation from matching. The agent may accomplish this in two basic ways. First, the agent can augment the condition by including as a required feature something missing from the current situation but present in reinforced examples. Alternatively, the agent can modify the condition to exclude situations manifesting some feature present in the counterexample but absent from training examples.

We can perform these operations consciously and deliberately using IM to identify criterial differences and including those as features in the engrams used for learning. NeuMAN performs such refinement implicitly through the strengthening of reinforced sequences augmented by the emergence of higher-level features that distinguish positive and negative examples. Learners can employ deliberate processes to identify features that guarantee appropriate responses or block inadequate conditions from triggering erroneous responses. IM makes it easy for the agent to scan memory for events manifesting all features of a conjectured situation.

7.6. Inferring a Rule

Whenever a set of reinforced responses reliably follows a set of antecedent conditions, NeuMAN stores and learns $S^*(t) \rightarrow S^*(t+1)$, essentially predicting that the common successor features will occur in response to the common antecedent features. IM produces this learning by automatically strengthening synapses from cells in $S^*(t)$ to cells in $S^*(t+1)$. Agents can also apply IM deliberately to conjecture candidate rules. As an example, in learning from failures, agents may seek to identify common preconditions they should avoid. In driving, for example, an agent might occasionally accidentally back into obstacles. To avoid such accidents, the agent would look for common antecedent conditions that occurred in each case and seek to block those in the future. Drivers may notice that each such case occurred when they failed to check one of the available sensors or views of the area behind the car. From that, this agent predicts that a failure to check precedes an accident, and this in turn enables it to learn that checking all sensors and views reliably predicts backing up without accident.

In general, when an agent can construct partial descriptions of hypothetical situations, it can search memory for patterned sequences to identify rules. Because NeuMAN employs extensional, non-symbolic representations, learning occurs continually without any sequential algorithms like that in the original formulation of IM [59].

7.7. Inferring a Procedure

With NeuMAN, the agent learns procedures the same way it learns hierarchical situation models. As an example, a language learner such as LaMDA [69] learns to generate responses to received sequences of words. Its models of word sequences correspond to the categories and patterns of syntax and semantics. Training reinforces LaMDA's best predictions of successions. When its predictions are connected to effectors, i.e. when it produces text responses, they become actions, and when actions follow from hierarchical compositions, we call those procedures.

The agent uses IM to learn hierarchical patterns from reinforced sequences. When learned pattern components produce actions, the agent exhibits procedure learning. Many researchers have pointed out that learning hierarchical procedures, also called plans, underlies much of human intelligence.[1,48,70]

8. PLANS AND THE STRUCTURE OF BEHAVIOR

Many scientists have described human plans as hierarchical compositions of a sequence of lower-level plans or atomic actions.[1, 39, 40] NeuMAN's continuous learning of reinforced patterns first produces low-level models and subsequently produces hierarchical compositions of reinforced sequences including component subpatterns. Some of the contained cells correspond to effector neurons that control muscles and generate observable behavior. Reinforcement of these behaviors strengthens the paths that produced them. In this way, a NeuMAN agent successively identifies and strengthens unitary and composite plans.

All of us are familiar with deliberate efforts to learn new skills, such as a foreign language, a physical skill, a song, a dance or use of a new tool. In each case, the most effective training starts by recruiting existing skill or knowledge components into a new sequential pattern that makes that pattern available as a new building block. We then compose new patterns using the highest-level available building blocks [71]. A skilled behaviorist will also encourage us to learn compound sequences from the end first, leaving the start of the to-be-learned sequence till last. Learning in that order means that we continually repeat and practice the earlier learned last parts of the sequence, and this repetition strengthens and automates those fragments. Automation of the remainder of the sequence reduces time to learning, because later elements of the series become unitized and automatic through repetition. The learner reduces the task of learning an entire sequence to one of learning how to trigger the learned pattern from the element that precedes it.

Skilled trainers know what capabilities the learning agents bring to the table. They provide training examples that employ those concepts and patterns as building blocks. The greater the knowledge and experience, the higher the level of components available.

People do not learn everything they are exposed to in their lives, for a number of reasons. First, most of the time, no external agents provide us reinforcement. Second, although repetition alone appears to reinforce behavior weakly, most reinforcement results from learned or secondary reinforcers.[72]. Further, as humans mature, the things that motivate and reinforce them change from mostly concrete to mostly abstract entities.[73] Finally, because animal brains develop and age over time, natural learners have critical periods for acquiring various capabilities and skills. If the agent misses the critical window, it may never master the corresponding skill.[74-77]

In summary, a NeuMAN agent learns to recognize patterns and to generate reinforced actions, but the agent may miss time-limited opportunities. Mastering the sounds of a language, the spatial awareness of a gymnast, and the swift and precise hand-eye control of a superb racquet player almost certainly requires early and continuous training. In a normal life, a limited range of reinforcement leads to a limited range of modeling and skill. If you did not learn a tonal or click language in childhood, you will probably never master it. If you did not learn to perceive rhythms and move rhythmically as a child, you likely will never feel talented in those ways.

9. EVERYTHING LEARNABLE IS LEARNED

Many important results in logic and computer science over the last 60 years inform the NeuMAN model proposed in this paper. In sum, they allow us to reject symbolic reasoning as a plausible foundation for automaticity, one of the essential characteristics of natural intelligent behavior. However, recent results have shown that sufficiently capable NNs, obviously automatic in generating responses, memorize everything reinforced and perform IM implicitly on all reinforced engrams. Because those engrams include sequences and learned patterns, NeuMAN reinforces all inferable relationships. Competition among alternative predictions based on diagnostic strength determines the agent's response to any situation. Thus, the HM basis of NeuMAN coupled with the reinforced IM inferences accomplishes the maximal amount of learning continually.

One finding about the holographic memory characteristics of sufficiently capable NNs stands out as key. These networks will learn everything learnable. Consider the basic learning problem of classifying stimuli as either members of some concept C or not C . In mathematics, we would say the agent must learn a function FC such that $FC(s) = \text{True}$ if and only if $s \in C$. Early work on NNs included the perceptron [78], a single-layer NN trained on such problems. Minsky and Paper [79] showed that the perceptron could only learn functions that linearly divided the feature space used to describe training stimuli. Perceptrons could learn only simple concepts, those with linearly

separable sets for positive and negative examples. For decades, this made many people skeptical about the utility of NNs.

We have referred to 21st century NNs as sufficiently capable, because they include a sufficient number of layers and wide dendritic trees between layers, as well as other features such as convolution, recurrence, and deep learning. These capabilities mean the NNs will learn every reinforced pattern they are exposed to. So NeuMAN will faithfully learn every consistently reinforced experience, including $s_k \in C$ and $s'_j \notin C$ for all examples s_k and counterexamples s'_j of C . The NNs record and exploit every possible training example appropriately.

The extensive finite storage of NNs means that the agent learns from every experience and does not need to reduce its knowledge to shorthand symbolic formulations that succinctly summarize training.

Humans constructed symbolic reasoning because it enabled them to apply explicit deductive procedures and reach provable conclusions. These results gave rise to most of the mathematical and logical research results for centuries. But this entire approach favors deductive logic over empirical inference, finding provable results using sequential procedures that can take significant amounts of time. In contrast, NeuMAN favors detailed explicit empirical sequences as a basis for inference. When the two approaches are faced with a novel test example, they behave differently. The symbolic learner will have formulated a functional description for the concept classification rule. That function will attempt to classify the test example using whatever logical rule it has inferred. Absent supervision, the agent's classification decision constitutes a guess as to the correct decision. A symbolic classifier will conjecture how best to generalize, meaning that under different scenarios it will guess correctly or incorrectly.

In contrast, when faced with an untrained example, NeuMAN will also make a guess based on the strongest, most diagnostic feature set present in the training example. If each feature set is considered a schema, NeuMAN uses every feature present in the stimulus to find every trained response, choosing the strongest response for its decision. So NeuMAN's guessing errors result from lack of knowledge. In contrast, the erroneous guesses of symbolic learners reflect an incorrect, overly general inferred classification rule.

All experiences are finite in the sense that engrams contain a finite number of active cells and the total number of successive engrams is finite. Furthermore, the sensory manifolds providing base inputs to the engrams describe the agent's environment thoroughly using finite extensional feature sets. When a computer represents relations by enumerating all possible values, we term those extensional representations. In contrast, when computers represent relations using symbolic variables associated with infinite domains, such as space and time, we term those intensional representations.

We conclude that the primary advantages of NeuMAN result from its powerful extensional situation models. NeuMAN learns every pattern inferable from a sequence of training examples, making errors only when training has been inconsistent, incomplete, or errorful. In short, we can effectively train NeuMAN to make no errors other than misclassifying a novel untrained example or an incorrect training instance. Later we point out that we can improve system trustworthiness by having it recognize and quickly adapt to such cases. Although current NN systems do not exhibit such adaptive interrupts, humans do and machines should.

10. CONCLUSIONS

This paper presents a new model for computational cognition that synthesizes a broad array of scientific findings in AI and cognitive science. Recent significant advances in NNs have unlocked a riddle that has persisted for decades: If humans reason symbolically, how does learning produce automated behaviors and how can the underlying machinery produce immediate responses to stimuli that match preconditions? The richly productive vein of symbolic reasoning, central to AI and cognitive science for most of the 20th century, fails at addressing that riddle.

The alternative we describe rests on recent discoveries that neural networks of appropriate design and adequate capabilities learn and apply models from copious reinforced training sequences. The NeuMAN computational model incorporates these findings. In addition, NeuMAN exploits learned patterns of predictable sequences to generate expectations that promote attention and response to high-value information. Roughly speaking, NeuMAN learns everything it trains on, acquires hierarchical models for sensing and acting within reference frames, and selects the strongest among competing alternative responses. Advanced NN applications exhibit many of NeuMAN's design features.

NeuMAN offers a mechanism to learn and make automatic complex behavior. It avoids the combinatorial delays of symbolic reasoning by relying on sensory manifolds to provide an extensional basis for modeling situations. Such situations stimulate, recall, and trigger learned responses. These capabilities will revolutionize computing and our understanding of computational intelligence. These capabilities suggest that NeuMAN provides a promising architectural framework for further advances in hardware and software. For many reasons we thus expect NeuMAN to catalyze significant advances across several fields.

REFERENCES

- [1] Miller, G.A., Galanter, E. and Pribram, K.H (1960). *Plans and the structure of behavior*. Henry Holt and Co. <https://doi.org/10.1037/10039-001>
- [2] Hudson, J. A., Fivush, R., & Kuebli, J. (1992). "Scripts and episodes: The development of event memory." *Applied Cognitive Psychology*, 6(6), 483-505.
- [3] Neisser, U. (2014). *Cognitive Psychology* (1st ed.). Psychology Press. doi.org/10.4324/9781315736174
- [4] Schank, R. C. (2014). *Conceptual information processing* (Vol. 3). Elsevier.
- [5] Hawkins, J. et al. (2019) "A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex." *Frontiers in Neural Circuits Journal*, 12, 121, Jan 11, 2019.
- [6] Hayes-Roth, F. (1977). "Uniform representations of structured patterns and an algorithm for the induction of contingency-response rules." *Information and control*, 33(2), 87-116.
- [7] Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability. A guide to the theory of NP-completeness*. New York: Freeman. <https://bohr.wlu.ca/hfan/cp412/references/ChapterOne.pdf>
- [8] Morin, F., & Bengio, Y. (2005). "Hierarchical probabilistic neural network language model." In *International workshop on artificial intelligence and statistics* (pp. 246-252). PMLR. <https://proceedings.mlr.press/r5/morin05a/morin05a.pdf>
- [9] Botvinick, M. M. (2008). "Hierarchical models of behavior and prefrontal function." *Trends in cognitive sciences*, 12(5), 201-208.
- [10] Chung, J., Ahn, S., & Bengio, Y. (2016). "Hierarchical multiscale recurrent neural networks." arXiv preprint arXiv:1609.01704.
- [11] Montavon, G., Samek, W., & Müller, K. R. (2018). "Methods for interpreting and understanding deep neural networks." *Digital signal processing*, 73, 1-15.
- [12] Ghorbani, A., Abid, A., & Zou, J. (2019). "Interpretation of neural networks is fragile." *Proceedings of the AAAI conference on artificial intelligence*, 33, 3681-3688.
- [13] Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. (2017). "Understanding deep learning requires re-thinking generalization." *ICLR 2017*. <https://openreview.net/pdf?id=Sy8gdB9xx>

- [14] Fischer, B. (1973). "Overlap of receptive field centers and representation of the visual field in the cat's optic tract." *Vision Research*, "(11), 2113-2120.
- [15] Burge, J., & Hayes-Roth, F. (1976). A novel pattern learning and classification procedure applied to the learning of vowels." *ICASSP'76. IEEE International Conference on Acoustics, Speech, and Signal Processing*, **1**, 154-157.
- [16] Sengupta, A., Pehlevan, C., Tepper, M., Genkin, A., Chklovskii, D. (2018) "Manifold-tiling localized receptive fields are optimal in similarity-preserving neural networks." In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, **31**, 7080–7090. Curran Associates.
- [17] Brown, T.B., *et al.* (2020). "Language Models are Few-Shot Learners." arXiv:2005.14165v4 (cs.CL) 22 Jul 2020.
- [18] Ramesh, A. (2022) *et al.* "Hierarchical Text-Conditional Image Generation with CLIP Latents." arXiv:2204.06125v1.
- [19] Vincent, J. (2022) The Verge. "An AI generated artwork's state fair victory fuels arguments over 'what art is'." <https://www.theverge.com/2022/9/1/23332684/ai-generated-artwork-wins-state-fair-competition-colorado>.
- [20] Jumper, J., Evans, R., Pritzel, A. *et al.* (2021) "Highly accurate protein structure prediction with AlphaFold." *Nature* **596**, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- [21] Merali, Z. (2022) "AlphaFold developers win US\$3-million Breakthrough Prize." *Nature*, 2022 Sep; 609(7929):889. doi: 10.1038/d41586-022-02999-9. PMID: 36138210..
- [22] Hayes-Roth, F. (1974). "Schematic classification problems and their solution." *Pattern Recognition*, **6**(2), 105-113.
- [23] Hayes-Roth, B., & Hayes-Roth, F. (1977). "The prominence of lexical information in memory representations of meaning." *Journal of Verbal Learning and Verbal Behavior*, **16**(1), 119-136.
- [24] Perez, E. (2016). "Deep learning machines are holographic memories." <https://medium.com/intuitionmachine/deep-learning-machines-are-holographic-memories-258272422995>
- [25] Reid, S., *et al.* (2022). "A Generalist Agent." arXiv:2205.06175v2 (cs.AI) 19 May 2022.
- [26] Costello, F. J., & Keane, M. T. (2001). "Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **27**(1), 255.
- [27] Hogendoorn, H. and Burkitt, A. N. (2019) "Predictive Coding with Neural Transmission Delays: A Real-Time Temporal Alignment Hypothesis." *eNeuro* **6**(2), 0412-18; DOI: 10.1523/ENEURO.0412-18.2019
- [28] Manning, C.D., Clark, K., Hewitt, J., Khandelwal, U. and Levy, O. (2020). "Emergent linguistic structure in artificial neural networks trained by self-supervision." *Proceedings of the National Academy of Sciences (PNAS)*, **117**(48), 30046-30054. <https://doi.org/10.1073/pnas.1907367117>
- [29] Hayes-Roth, F. (2006a). "Model-based communication networks and VIRT: Orders of magnitude better for information superiority." In *MILCOM 2006-2006 IEEE Military Communications Conference*, (pp. 1-7). IEEE.
- [30] Denning, P. J. (2006). "Infoglut." *Communications of the ACM*, **49**(7), 15-19.
- [31] Hayes-Roth, F. (2006b). "Valued Information at the Right Time (VIRT): Why less volume is more value in hastily formed networks." Naval Postgraduate School, Monterey, CA.
- [32] de-Wit, L., Machilsen, B., and Putzeys, T. (2010) *Predictive coding and the neural response to predictable stimuli. Journal of Neuroscience*, **30**(26), 8702-8703. DOI: <https://doi.org/10.1523/JNEUROSCI.2248-10.2010>
- [33] Rao, R. P., & Ballard, D. H. (1999). "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects." *Nature Neuroscience*, **2**(1), 79-87.
- [34] Huang, Y., & Rao, R. P. (2011). "Predictive coding." *Wiley Interdisciplinary Reviews: Cognitive Science*, **2**(5), 580-593.
- [35] Battistoni E., Stein, T., Peelen, M.V. (2017) "Preparatory attention in visual cortex." *Ann. N Y Acad. Sci.* **1396**:92–107. doi:10.1111/nyas.13320 pmid:28253445
- [36] van Moorselaar, D., Lampers, E., Cordesius, E., & Slagter, H. A. (2020). "Neural mechanisms underlying expectation-dependent inhibition of distracting information." *Elife*, **9**, e61048.
- [37] Silver, D., Singh, S., Precup, D., Sutton, R. S. (2021) "Reward is Enough." *Artificial Intelligence*, 229, <https://doi.org/10.1016/j.artint.2021.103535>
- [38] Wesson, R. B. (1977). "Planning in the World of the Air Traffic Controller." In *IJCAI*, **5**, 473-479.

- [39] Hayes-Roth, F., L. D. Erman, A. Terry, B. Hayes-Roth (1992). "Distributed Intelligent Control and Management: Concepts, methods and tools for developing DICAM applications." *SEKE*: 235-244.
- [40] Albus, J. S., & Rippey, W. G. (1994). "RCS: A reference model architecture for intelligent control." *Proceedings of PerAc'94. From Perception to Action*, 218-229. IEEE.
- [41] Miao, A. X., Zacharias, G. L., & Kao, S. P. (1997). "A computational situation assessment model for nuclear power plant operations." *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, **27**(6), 728-742.
- [42] Scheel, O., Schwarz, L., Navab, N., & Tombari, F. (2018). "Situation assessment for planning lane changes: Combining recurrent models and prediction." *IEEE International Conference on Robotics and Automation (ICRA)*, 2082-2088.
- [43] Miller, G.A., Galanter, E. and Pribram, K.H (1960). *Plans and the structure of behavior*. Henry Holt and Co. <https://doi.org/10.1037/10039-001>
- [44] Hayes-Roth, B., & Hayes-Roth, F. (1979). "A cognitive model of planning." *Cognitive Science*, **3**(4), 275-310.
- [45] Johnson-Laird, P. N. (1980). "Mental models in cognitive science." *Cognitive Science*, **4**(1), 71-115.
- [46] Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.
- [47] Minsky, M. (1988). *Society of mind*. Simon and Schuster.
- [48] Albus, J. S., & Meystel, A. (2001). "Engineering of mind: An introduction to the science of intelligent systems." NISR.
- [49] Eppe, M., Gumbsch, C., Kerzel, M., Nguyen, P. D., Butz, M. V., & Wermter, S. (2022). "Intelligent problem-solving as integrated hierarchical reinforcement learning." *Nature Machine Intelligence*, **4**(1), 11-20.
- [50] Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.
- [51] Pribram, K. H. (1986). "The cognitive revolution and mind/brain issues." *American Psychologist*, **41**(5), 507.
- [52] Cavanagh, J. P. (1975). "Two classes of holographic processes realizable in the neural realm." In *Formal aspects of cognitive processes*, 14-40. Springer, Berlin, Heidelberg.
- [53] Hayes-Roth, F., & Mostow, D. (1975). "An automatically compilable recognition network for structured patterns." *IJCAI*, **4**, 356-362.
- [54] Mostow, D. J., & Hayes-Roth, F. (1978). "A production system for speech understanding." In *Pattern-Directed Inference Systems*, 471-481. Academic Press.
- [55] Hayes-Roth, B. (1977) "Evolution of cognitive structure and processes." *Psychological Review*, **84**(3), 260.
- [56] Holland, J. H. (1992). "Genetic algorithms." *Scientific American*, **267**(1), 66-73.
- [57] Onen, M. et al., (2022). "Nanosecond protonic programmable protonic resistors for analog deep learning." *Science*, **377**(6605), 539-543.
- [58] DeBole, Michael V., et al. (2019) "TrueNorth: Accelerating from zero to 64 million neurons in 10 years." *Computer*, **52**(5), 20-29.
- [59] Hayes-Roth, F., & McDermott, J. (1978). "An interference matching technique for inducing abstractions." *Communications of the ACM*, **21**(5), 401-411.
- [60] Oaksford, M., & Chater, N. (2017). "Causal models and conditional reasoning." *The Oxford handbook of causal reasoning*, 327ff.
- [61] Hayes-Roth, F. (1983). "Using proofs and refutations to learn from experience." *Machine Learning*, 221-240. Springer, Berlin, Heidelberg.
- [62] Hayes-Roth, F. (1989) "Towards benchmarks for knowledge systems and their implications for data engineering." *IEEE Transactions on Knowledge & Data Engineering*, **1**(1), 101-110.
- [63] Pease, A., Smaill, A., et al.. (2010). "Applying Lakatos-style reasoning to AI problems." *Thinking Machines and the philosophy of computer science: Concepts and principles*, 149-174.
- [64] Vere, S. A. (1980). "Multilevel counterfactuals for generalizations of relational concepts and productions." *Artificial Intelligence*, **14**(2), 139-164.
- [65] Hayes-Roth, F., Klahr, P., & Mostow, D. J. (1980). "Advice-taking and knowledge refinement: An iterative view of skill acquisition." Rand Corp., Santa Monica, CA.
- [66] Quinlan, J. R. (1986). "Induction of decision trees." *Machine Learning*, **1**(1), 81-106.
- [67] Mitchell, T. M. (1977). "Version spaces: A candidate elimination approach to rule learning." In *Proceedings of the 5th international joint conference on Artificial intelligence*, 305-310.

- [68] Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). *Machine learning: An artificial intelligence approach*. Springer Science & Business Media.
- [69] Thoppilan, R., De Freitas, *et al.* (2022). “LaMDA: Language models for dialog applications.” arXiv preprint arXiv:2201.08239.
- [70] Byrne, R. W., & Russon, A. E. (1998). “Learning by imitation: A hierarchical approach.” *Behavioral and Brain Sciences*, **21**(5), 667-684.
- [71] Dyson, G. B. (2012). *Darwin among the machines: The evolution of global intelligence*. Basic Books.
- [72] Pribram, K. H. (2013). “The enigma of reinforcement.” *Neurobehavioral Plasticity*, 399-422. Psychology Press.
- [73] DeKeyser, R. M. (2000). “The robustness of critical period effects in second language acquisition.” *Studies in second language acquisition*, **22**(4), 499-533.
- [74] Hensch, T. K. (2004). “Critical period regulation.” *Annu. Rev. Neurosci.*, **27**, 549-579.
- [75] Hensch, T. K. (2005). “Critical period plasticity in local cortical circuits.” *Nature Reviews Neuroscience*, **6**(11), 877-888.
- [76] Kral, A. (2013). “Auditory critical periods: a review from a system’s perspective.” *Neuroscience*, **247**, 117-133.
- [77] Rosenblatt, F. (1958) “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.” *Psychological Review*, **65**, 386. <https://doi.org/10.1037/h0042519>
- [78] Minsky, M., & Papert, S. A. (2017). *Perceptrons, Reissue of the 1988 Expanded Edition with a new foreword by Léon Bottou: An Introduction to Computational Geometry*. MIT press.

ACKNOWLEDGEMENTS

Neil Jacobstein provided a thoughtful review of an earlier draft as well as many relevant observations.

AUTHOR

Frederick (“Rick”) Roth retired as a Professor of Information Sciences at the Naval Postgraduate School in Monterey, California. He is a Fellow of the Association for the Advancement of Artificial Intelligence, formerly CTO/Software for Hewlett Packard, and founder and CEO of several companies, including Cimflex Teknowledge, a public AI and robotics business. He formerly published under the surname Hayes-Roth.

