Emulating a Brain System

Kenneth M'Balé and Darsana Josyula
Department of Computer Science
Bowie State University
Bowie, MD 20715 USA
kmbale@cs.umd.edu, darsana@cs.umd.edu

Abstract

Can intelligence be produced simply by reverse engineering the brain of an intelligent animal? In this paper, we argue that such reverse engineering will be ineffective if the focus is on reverse-engineering the brain mapping. We believe that the brain "hardware" implements a "system" and our work focuses on emulating this basic brain system and to provide it the necessary interfaces to support a collective intelligence.

Introduction

Can brain-mapping data be used to reverse engineer a brain system *in silico*? This is actually the question of whether consciousness is fully contained within the physical structure that is the brain. Do the brain and its supporting systems fully account for consciousness or are there other components that transcend the body that are also at play? If metaphysical components play a role, then the answer is negative, since mapping just the anatomical aspects of the consciousness system would leave a critical component missing and, hence will not suffice to produce consciousness. A related question is whether consciousness is the result of a static system operating in the brain, or, is it a function of a dynamic system? In other words, what if consciousness is not a function of the state of discrete anatomical objects in the brain, synapses and neurons, but a function of the exchange of signals and chemicals between them? In such a case, the challenge is to replicate their function and interaction rather than their state. Can such a dynamic system be reverse engineered and captured? A similar question would be; can we replicate the Internet by capturing the state of every connected computer at a moment in time? What makes the Internet the useful tool it has become is the information flow between connected computers, not the specific state of discrete nodes on the network. The next challenge is the question of the subconscious. From psychology, we are aware of its existence and its impact on our lives. However, our understanding of the subconscious is severely limited. How can we determine if we have effectively reverse engineered the subconscious since we cannot directly test its effects?

Related Work

Noam Chomsky discusses the evolution of the field of artificial intelligence from 1956, when John McCarthy defined the science, until today (Ramsay, 2012). The goal of AI was to study intelligence by implementing its essential features using man-made technology. This goal has resulted in several practical applications people use every day. The field has produced significant advances in search engines, data mining, speech recognition, image processing, and expert systems, to name a few.

The engineering of these practical solutions has taken AI in a direction that enables the rapid implementation of the essential features of intelligence these applications require. A search engine can be very efficient at finding relevant results but it does not comprehend what it is searching for. A data mining application can identify relevant features from noise in a dataset, but it does not comprehend the meaning or significance of what it finds.

To fill the void created by the absence of comprehension, AI researchers rely on formalisms and, more recently, on statistical methods. Modern AI has abandoned the use of formalisms (Eiter & Simkus, 2007), in favor of probabilistic and statistical models (Darwiche, 2010), in its decision-making. This shift reflects the substantial increase in computing capacity to process Big Data.

Statistical methods are effective at identifying salient features and at predicting the next event. However, they neither impart nor proceed from comprehension. Comprehension requires intelligence. These applications are tools in the hand of the intelligence that still resides within the user.

A consequence of statistical methods is a loss of transparency. Often, the processing of these applications is difficult for a user to understand, even though the results are understandable. We view this lack of transparency as a reflection of our lack of understanding of how intelligence works within our own biological hardware. Of course, there is no assumption that our brains use statistical methods to achieve cognition. The statistical methods achieve sufficient approximation of intelligence to be useful within the narrow application domain.

Other approaches focus on emulating the biological architecture of the brain. These approaches are based on the hypothesis that the brain is a collection of simple systems that collaborate to produce intelligence, and it makes sense to emulate this architecture to produce the same result. This line of thinking has resulted in several thrusts such as neural networks, and more recently, *Connectomics (Kelland, 2011)*. Connectomics aims to map the brain's synapses in an effort to decipher how information flows through the brain's neural circuits. From the standpoint of neuroscience, the goal is to understand how the brain generates thoughts and perceptions. While this area of research will undoubtedly yield positive results in our struggle against diseases such as dementia and schizophrenia, it is not clear how it can provide insight into how intelligence works.

In (Sejnowski, 2012), the author wonders; *if experiences are coded into brain connections, could a wiring diagram simulate your mind?* Even if such a simulation happens, our understanding of intelligence would not have significantly advanced. In (Brenner & Sejnowski, 2011), the authors comment that *this is a good time to pause and ask ourselves what we expect to find at the end of this immense omic brainbow*. Brenner is largely credited with establishing brain mapping but he does not believe this path will yield results for our understanding of cognition.

The feasibility of observing the brain in action is still in question (Underwood, 2014). The question is whether the functioning of the brain, observed at its core level, will make sense to the researcher. This can only happen if the reality of the systems of the brain is a subset of human reality. Otherwise, the researcher will not have a frame of reference to understand intelligence even if she can replicate it. Can a researcher understand what drives an animal by taking it apart? Which chemical test does she use to determine whether a cat enjoys tickles? The frame of reference appears mismatched; advanced intelligence is not a direct function of the anatomical hardware it operates upon.

Our current limited understanding of the brain shows that various regions of the brain are dedicated to perform certain functions, including enabling communications between regions. In (Kak, Donald, & Delaune, 2005), the authors conclude that *the evidence from neuroscience that we reviewed showed how specific centers in the brain are dedicated to different cognitive tasks. But these centers do not merely do signal processing: each operates within the universe of its experience so that it is able to generalize individually. This generalization keeps up with new experience and is further related to other cognitive processes in the brain. It is in this manner that cognitive ability is holistic and irreducible to a mechanistic computing algorithm. Viewed differently, each agent is an apparatus that taps into the "universal field of consciousness." On the other hand, AI machines based on classical computing principles have a fixed universe of discourse so they are unable to adapt in a flexible manner to a changing universe. This is why they cannot match biological intelligence.*

Reference Architecture

Our work views the conscious brain as a collection of systems that can adapt flexibly to specialize or generalize based on the need. In this section, we discuss the reference architecture that our artificial agent is based upon.

Component

The basic component of the reference architecture accepts a stimulus, confirms that it matches certain criteria, and if it does, produces a response. See Figure 1.

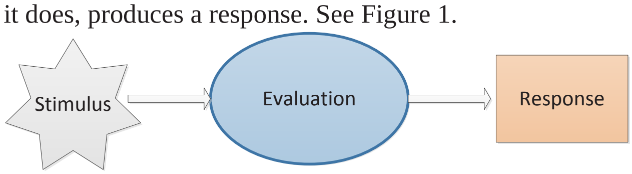

Figure 1 - Stimulus Response Model

Neurons implement this component. In other words, it is possible to identify a specific set of neurons that implement this component for a given stimulus-response. The configuration of the neurons implements the component without hardwiring.

The component varies in four dimensions of complexity; stimulus, evaluation, response and time. All permutations are valid. For example, a complex stimulus might involve several stimuli. A complex evaluation might involve several combinations of weights and strengths. A complex response might involve several actions or steps. A component might operate over a brief or long period.

As depicted on Figure 2, we can now differentiate between two component types, which we call basic and derived. The component described above is a basic component. A derived component accepts another component's response as its stimulus.

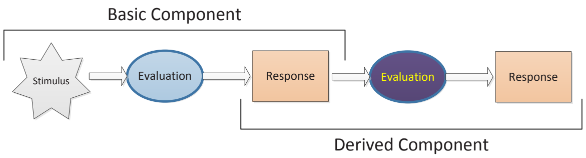

Figure 2 Basic and Derived Components

System

A component combines with other basic and derived components to form a system. We can envision that, given recurring stimuli that trigger an existing system, a cognitive system seeks to simplify the evaluation and response in an effort to produce the final response as soon as possible. This strategy is consistent with Evolution that favors rapid, almost reflex speed reactions to important stimuli. The evolutionary advantage of a sophisticated cognitive system is that it can create complex systems in response to complex changes in the environment. The cost is that it is undesirably slow when placed in the critical path to the response. Therefore, the cognitive system needs to learn from the environment which stimuli require which favorable response. Until the proper system is created, it makes sense to invest the time and effort in the cognitive system. Once the proper system is created, the cognitive system no longer needs to be engaged until unexpected results occur.

In this model, the brain generates and operates systems to produce a response to stimuli. There are several systems, overlapping in stimuli and responses. What selects the proper system? This selector may be the intelligence that we seek to replicate.

Referring to Figure 3, we envision a *mechanism* as a special system that has a control interface to anatomical elements. Thus, a mechanism is a hardwired system of the brain that is connected to senses, limbs, organs and other anatomical elements of the body. A mechanism is able to control the anatomical element and receives information about the state of the element. In addition, a mechanism can communicate with another system.

Consider a basketball player shooting to score. A mechanism moves the hands and arms to shoot the ball. Another mechanism moves her feet and jumps. Another mechanism orients her head and eyes towards the basket. When the player is just learning the game, the cognitive system is actively involved and, as a result, the shooting action is not smooth or natural. With enough practice, the "shoot the basket" system is created and the shooting action becomes second nature. It no longer requires the cognitive system.

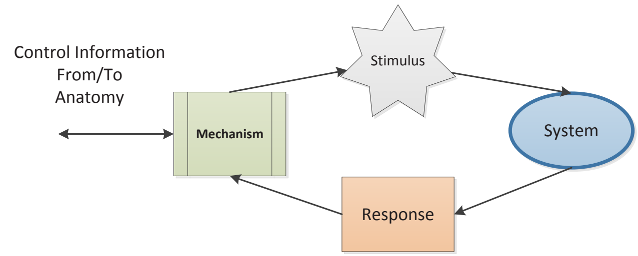

Figure 3 Systems and Mechanisms

Intelligence as Coordination of Systems

In the static behaviorism model of Figure 4, a system coordinates the mechanisms (and other systems if necessary).

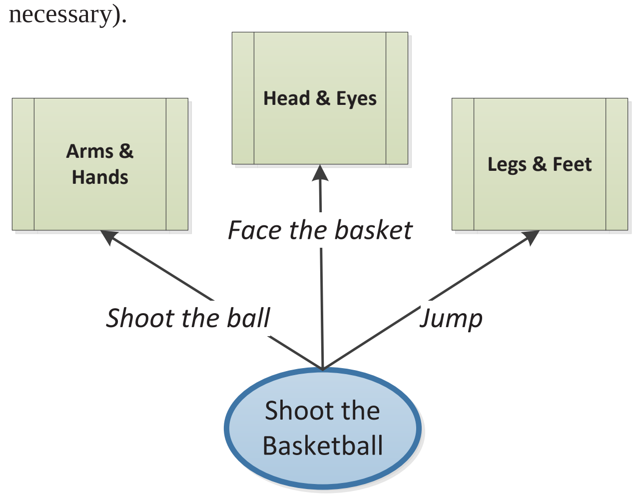

Figure 4 Shoot the Basketball (Static Behaviorism)

In the dynamic emergence model of Figure 5, systems and mechanisms coordinate with each other. In the static behaviorism model, the coordinator system maps the objective (shoot the basketball) to inputs to mechanisms and stimuli to other systems. We can imagine that until it learns which systems and mechanisms are actually required to accomplish this objective, it could involve unnecessary ones. For example, it might involve the "Nose and Tongue" mechanism to "smell and taste the ball." In the dynamic emergence model, all systems know the objective and act to achieve it in the manner they innately can. The "Nose and Tongue" mechanism "smells and tastes the ball" because that is what it is capable of. Every system (and mechanism) reacts in the way it knows how. This characteristic explains why even a small change can throw off the basketball player. For example, if the ball smells funny or feels wrong, the cognitive system gets engaged again in response to the anomaly.

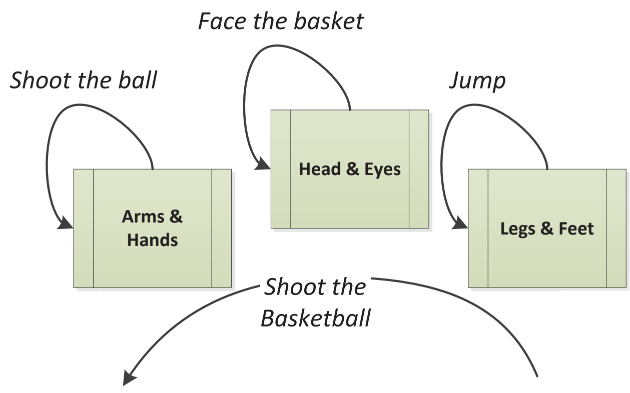

Figure 5 Shoot the Basketball (Dynamic Emergence)

It is not clear which model is correct since the brain is an adaptive organ. Some researchers point out that the brain is able to perform up to a point even when damaged. They use this observation to invalidate the notion of a designated coordination system. On the other hand, perhaps the coordination system moves elsewhere. In other words, similar to other systems in the body, the systems are generic and could specialize based on the need. There is always a coordination system but any system can become the coordinator. This is similar to a hive losing its queen.

The important observation is that once the system is in place, it becomes virtually hardwired and capable of performance similar to the inherited systems. As a result, it realizes its evolutionary benefit of enabling adaptation and therefore survival. It also appears that repetition and strength of stimuli are the key factors in creating a virtual hardwired system.

Anomalies

Note that an objective also defines an expectation. An expectation is an observable result from an action. Once the actions to shoot the basketball are performed, the expectation is that the basketball travels through the basket. Given an objective, each system has one or more expectations that result from its actions. In nature, it is not necessary for all aspects of the objective to have been met, for the intelligence to conclude that all aspects of the expectation have been met. The intelligence seeks to achieve equivalence between objective and expectation, but not identity. We refer to an anomaly as the condition when sufficient expectations have failed to materialize to conclude that the objective has been met.

While we have identified no physical difference between mechanisms and systems, we can organize them in terms of the flow of anomaly (a meta-stimulus) and response to the anomaly, as shown on Figure 6. A mechanism engages an inherited system to address its anomaly. The inherited system can then trigger other primary or derived systems to respond to the anomaly. If the inherited system encounters an anomaly of its own, it engages the cognitive system. Finally, if the cognitive system encounters an anomaly of its own, it engages a metacognitive system.

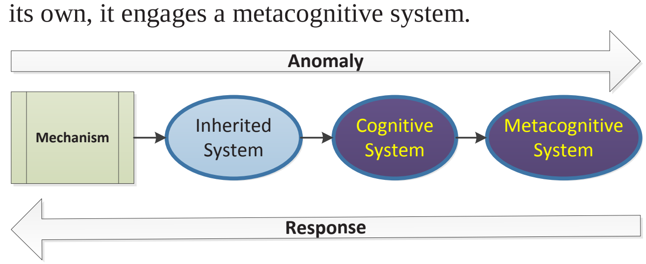

Figure 6 Anomaly Response

Collective Intelligence

A collective intelligence is a loose collaboration of systems with low or simple intelligence that produce results of high or complex intelligence. The collaboration is termed loose because there is no predetermined leader or coordinator. The individuals organize themselves in accordance with inherent rules and capabilities they possess. The methods and algorithms these individuals use to organize themselves to accomplish an objective are collectively referred to as swarm intelligence.

When artificially implemented, such as software or hardware, an individual system is called an agent. Typically, an agent is built with a specific objective in mind. For example, an agent can be designed to mimic a bird in a flock or a human being in a crowd or a molecule in a solution. An agent is universal when it is not bound to any specific objective. The universal agent must cope with an unbounded set of problems, with little inherent knowledge of the domain. The user determines the objective(s) the universal agent pursues.

Our architecture is based on the hypothesis that the reverse engineering of the physical brain is not necessary to replicate human level intelligence. Instead, we only need to reverse engineer a universal system. We can then instantiate a metacognitive system, a cognitive system and the inherited systems necessary to interact with the environment.

GPME

We call our universal agent the General Purpose Metacognition Engine (GPME) ((M'Balé & Josyula, 2014, 2013)). The GPME is what provides intelligent behavior to the brain system of our reference architecture. It is designed to interact only with a system or with another GPME contained within another system. Referring to Figure 7, we refer to the system that contains the GPME as the host. The host operates in an environment. The environment can be outside of or inside the body. To carry out its function, the host possesses cognitive capabilities specific to its purpose and function. The host senses stimuli from the environment and reacts. It uses actuators to affect the environment. The host provides some (coordinated model) or all (self-coordinated model) of the sensory information to the GPME as an observation. When the GPME detects an anomaly, it provides a tailored suggestion to the host. The host is not obligated to act on the suggestion. We refer to the communication between the GPME and the host as telemetry.

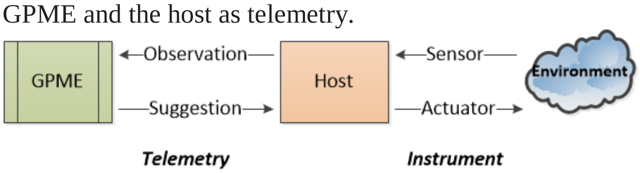

Figure 7 Context Diagram

Like the visual cortex of the brain, the host is a sophisticated system capable of cognitive functions autonomously. For example, assume the host is a robot capable of movement, equipped with a gripping arm, and auditory and visual sensors. The host has the ability to safely navigate a space from an original location to a target location. The GPME does not provide detailed step-by-step instruction to navigate from point A to point B. The GPME suggests the host to move from point A to point B. The host is sophisticated enough to act on this suggestion and report its status back to the GPME. We refer to it as a suggestion because the host may not be able to act or may not succeed in the act. While this example used a robot, it could also have used the brain system that manages vision.

Perpetual Cognition Circuit

The core design of the GPME operates a continuous cycle we call the Perpetual Cognition Circuit (See Figure 8). The GPME receives telemetry from the environment and from the host. The instruments create an observation. The observation triggers the learning apparatus to process the new observation into the knowledge base. The assimilation of a new observation changes the organization of the episodic memory. The projection apparatus uses the knowledge base to project the future wellbeing of the system. If the future wellbeing of the system is in jeopardy, it suggests actions that maximizes wellbeing and monitors success. The measure of wellbeing is called *homeostasis*. The purpose and only goal of the GPME is to maximize homeostasis. All other goals are effectively steps towards this overarching goal.

The GPME communicate with each other to form a collective intelligence. We refer to a group of communicating GPME instances as a GPME swarm. The GPME swarm enables its members to share their knowledge bases, accelerate learning and facilitate the construction of new brain systems.

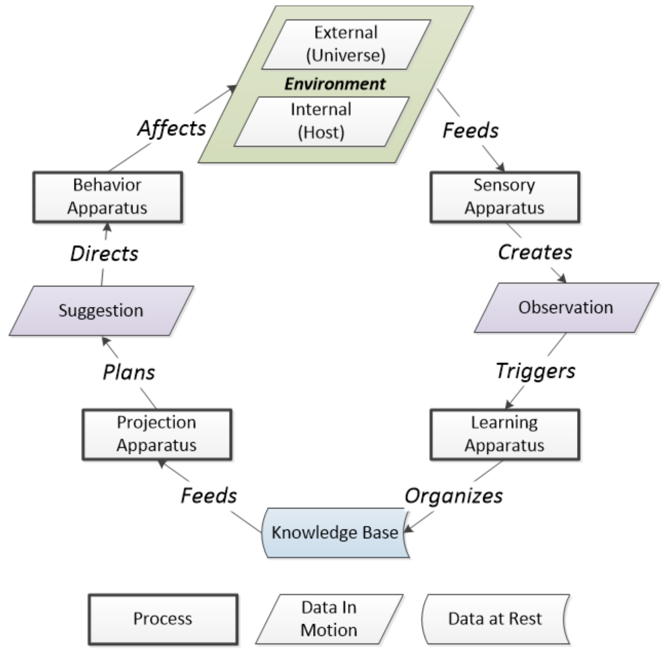

Figure 8 Perpetual Cognition Circuit

Learning

The GPME uses two learning methods; *progressive reduction* and *selective imitation*.

Progressive reduction applies observational learning techniques on the telemetry. Since the telemetry contains a great deal of noise, the GPME looks for special patterns we call *rhythmic* patterns. A rhythmic pattern exists when the GPME detects at least two correlated patterns in the telemetry. In order to detect correlated patterns, the GPME sections the telemetry into episodes. It then clusters episodes with similar correlated patterns to derive a case. Therefore, progressive reduction enables the GPME to generate cases and apply case based reasoning techniques to resolve anomalies. The resolution of anomalies involves selecting the case with the most positive impact on homeostasis. The selected case provides the information necessary to provide the host with a suggestion in response to the anomaly. The GPME then observes the results in the telemetry to adjust the utility rating of cases.

The GPME is an intrinsic reinforcement learner. It uses an internal reward that is a function of the number and types of anomalies that currently exist and the projection of the reward in the future. We refer to the result of this function as the *homeostasis*. As a result, the GPME can react to an anomaly caused by future unexpected homeostasis values.

How does a GPME detect an anomaly without inherent knowledge? An anomaly occurs when an expectation is not met. The GPME detects physical anomalies, such as the absence of stimuli, and logical anomalies, such as the absence of a result. As the GPME processes the telemetry, it projects forward the future state of the telemetry based on its case history. An anomaly is a substantial difference between the actual telemetry and the projected telemetry. For example, the GPME expects a particular stimulus to occur in future moment M_x based on the past. When moment M_x arrives, the stimulus is either present or absent. If it is absent, an anomaly has occurred.

In practice, the GPME implements this concept using a *bandwidth*. The bandwidth is a projected minimum and maximum for the stimulus value or occurrence. An anomaly exists when the observation is outside the projected band.

In addition to being an intrinsic reinforcement learner, the GPME learns by imitation. The formalization of the knowledge base makes it possible for the GPME to share parts of its knowledge base with other instances. The recipient GPME can then select the portions of the model's knowledge base that it wants to incorporate within its own. The GPME is able to learn new cases and reasoning mechanisms from other GPME instances without needing to experience the environment first hand.

Selective imitation involves receiving case information from other GPME. We refer to the imitation as selective because the GPME must determine whether to incorporate the case within its own knowledge base. All GPME in a swarm share all of their cases. It is up to the recipient to select which cases to accept. The selection process is based on three factors, host similarity, GPME maturity and mentor confidence.

The attributes and function of each host is described in an XML interface specification. The GPME's share their respective hosts' descriptions to determine whether there exists sufficient commonality to use each other's cases. The maturity of the source GPME is a function of the number of mature cases in its knowledge base. A GPME with a higher maturity level is less likely to select cases from a lower maturity GPME. Once a case is accepted, the GPME will treat it as any generated case. The case can be an example of what to do or an example of what not to do, based on the actual homeostasis impact. As a GPME (the learner) applies the cases acquired from another GPME (the mentor), the learner rates mentor's cases to derive a confidence level. The higher the confidence level the more likely the learner is to use cases it learned from the particular mentor.

Time and Decay

The GPME processes a great deal of raw information to generate cases. In addition, the GPME receives cases from mentors. To manage the volume of data, the GPME uses the concept of forgetting through decay. Each time a component of the GPME knowledge base is used, the GPME determines whether to increase or decrease its utility value. We refer to the utility value as the *Damaru*. An independent process within the GPME visits every knowledge base object and decrements the Damaru. The time it takes to visit every object once is an atomic GPME time unit we call a *moment*. Therefore, every moment, the Damaru is adjusted up or down. When it reaches within a predetermined range, the object is purged from the knowledge base.

Conclusion

Our hypothesis is that the creation of a brain system does not require the reverse engineering of anatomical component of the brain. We believe that the brain "hardware" implements a system of simple design that operates on an electro-chemical machine. Therefore, it is sufficient to emulate this basic brain system and to provide it the necessary interfaces to support a collective intelligence. The GPME is a universal agent that provides sufficient generality and capability to be the building block of a brain system.

Acknowledgments

This research is supported in part by the Office of Naval Research grant ONR #N00014-12-1-0430.

References

Brenner, S., & Sejnowski, T. J. (2011). Understanding the human brain. *Science (New York, N.Y.)*, *334*(6056), 567. doi:10.1126/science.1215674

Darwiche, A. (2010). Bayesian networks. *Communications of the ACM*, *53*(12), 80. doi:10.1145/1859204.1859227

Eiter, T., & Simkus, M. (2007). FDNC : Decidable Nonmonotonic Disjunctive Logic Programs with Function Symbols, *9*(9), 1–45.

Kelland, K. (2011). Scientists find way to map brain's complexity. *Reuters*. Retrieved from http://www.reuters.com/article/2011/04/10/us-brain-model-idUSTRE7392KU20110410

M'Balé, K. M., & Josyula, D. (2014). Handling Seasonality using Metacognition. In *Cognitive 2014*. Venice, Italy.

M'Balé, K. M., & Josyula, D. P. (2013). Integrating Metacognition into Artificial Agents. In *AAAI 2013 Fall Symposium Series* (pp. 55–62). Arlington, VA: AAAI Press.

Ramsay, G. G. (2012). Noam Chomsky on Where Artificial Intelligence Went Wrong. *The Atlantic*, 1–20.

Sejnowski, T. J. (2012, November 4). The product of our neurons. *New Scientist*, *334*(6056), 1. Retrieved from http://papers.cnl.salk.edu/PDFs/The product of our neurons 2012-4271.pdf

Underwood, E. (2014). BRAIN project meets physics. *Science*, *344*(6187), 954–955. doi:10.1126/science.344.6187.954