ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/361968889

Where is the mind within the brain? Transient selection of subnetworks by metabotropic receptors and G protein-gated ion channels

Preprint · July 2022

DOI: 10.13140/RG.2.2.21067.87843

CITATIONS
0

READS
441

1 author:

Danko Nikolić
(Ex) Max Planck Institute for Brain Research
114 PUBLICATIONS **4,782** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Test empirical predictions of the theory of practopoiesis View project

DIGITAL SYNESTHESIA View project

All content following this page was uploaded by Danko Nikolić on 13 July 2022.

The user has requested enhancement of the downloaded file.

Where is the mind within the brain? Transient selection of subnetworks by metabotropic receptors and G protein-gated ion channels

Danko Nikolić*

*Frankfurt Institute for Advanced Studies
*evocenta GmbH
*Robots Go Mental UG

Correspondence:

Dr. Danko Nikolić
Frankfurt Institute for Advanced Studies
Ruth-Moufang-Straße 1,
60438 Frankfurt am Main
Germany
danko.nikolic@gmail.com

Abstract

Perhaps the most important question posed by brain research is: How the brain gives rise to the mind. To answer this question, we have primarily relied on the connectionist paradigm: The brain's entire knowledge and thinking skills are thought to be stored in the connections; and the mental operations are executed by network computations. I propose here an alternative paradigm: Our knowledge and skills are stored in metabotropic receptors (MRs) and the G protein-gated ion channels (GPGICs). Here, mental operations are assumed to be executed by the functions of MRs and GPGICs. As GPGICs have the capacity to close or open branches of dendritic trees and axon terminals, their states transiently re-route neural activity throughout the nervous system. First, MRs detect ligands that signal the need to activate GPGICs. Next, GPGICs transiently selects a subnetwork within the brain. The process of selecting this new subnetwork is what constitutes a mental operation – be it in a form of directed attention, perception of a grandmother or making a decision. Synaptic connections and network computations play only a secondary role, supporting MRs and GPGICs. According to this new paradigm, the mind emerges within the brain as the function of MRs and GPGICs whose primary function is to continually select the pathways over which neural activity will be allowed to pass.

1. Introduction

Today, there is practically a unanimous agreement that the computations of the brain are achieved through network connectivity *(1)*. A variety of theories exist (e.g., 2,3,4) but they all propose different ways of how network computations may take place; all these theories agree that the synaptic connectivity contains the bulk the knowledge learned by the brain. The existing theories also agree that central to mental operations are the voltages forming across neuron membranes which carry the relevant information and computation. Consequently, it is believed that, at any moment in time, the current contents of a mind are determined by the collective states of the neuron membranes i.e., by their momentary voltages (1,2,3,4). Thus, these two mechanisms, synaptic connections and voltages, have been relied on as the key explanatory tools for all our attempts to understand how a mind may emerge within the brain. This paradigm is generally known as connectionism (1).

However, connectionism did not yet produce a satisfactory explanation of how the mental emerges from the physical. A number of open problems remains *(* 5,6,7,8*)*. As a result, the explanatory gap between the mind and the brain remains wide open (9,10,11,12). This situation raises a question of whether the classical paradigm should be challenged and an alternative paradigm should be proposed—one that can hopefully offer a fresh set of ideas on how to explain the mind within the brain.

2. An alternative paradigm: Transient selection of subnetworks by metabotropic receptors and G protein-gated ion channels

The idea proposed here is that, at any moment in time, only a fraction of the entire physical neural network is permeable for neural activity: a large portion of dendritic trees and terminal branches are usually closed for traffic (are not effectively connected) and thus, do not participate in the flow of neural activity (or the ability to participate in the flow is severely reduced) (Figure 1A, B). Furthermore, the decisions on which pathways will be permeable and which will not, are made by metabotropic receptors (MRs; aka, G protein-coupled receptors) and G protein-gated ion channels (GPGICs) (Figure 1C, D). These

membrane proteins are assumed to solely give rise to cognition by selecting subnetworks as they act as gates: they close and open the membranes for passage of voltage signals. Here, the job of an MR is to detect a condition under which a change should be made, and the job of GPGIC is to make that change. An MR detects a presence of a ligand in the extracellular space and this ligand is presumed to serve as a signal for the change. The presence of the ligand is signaled further to the GPGIC by releasing a G protein intracellularly. G protein then activates GPGIC, either directly by binding on the GPGIC, or indirectly via phosphorylation (13,14,15). The activation of a GPGIC alters the electric properties of the membrane (16,17) (Figures 1C, D)(a more detailed example of a contribution of a GPGICs is provided in the Supplementary Materials). The act of cognition (attention, perception decision making etc.) is the act of changing the electric properties of the membrane. This occurs as MRs and GPGICs distributed across a network select a subnetwork that will be functional (Figure 1E). By each selection, a new, unique sensory-motor loop is created (Figure 1F, G). Voltages and synapses do not do anything but such simple sensory motor loops; all the 'heavy' work is done by MRs and GPGICs. These loops exist only transiently because new MR and GPGIC changes are being made; and even if no changes would be made, the states of GPGICs spontaneously return back to their initial (default) states of inactivation (Figure 2).

This MR-GPGIC system is flexible for learning new skills because there are at least 50 types of MRs, making it possible to detect virtually every relevant type of ligand (18). Many of these ligands are neurotransmitters originating from nearby synapses. In addition, there are many types of GPGICs (19), specialized for various types of ions and producing various effects on the membrane's ability to conduct electric signals (13,14). MRs and GPGICs can be found in the central (20) and peripheral nervous system (21). This all allows enough flexibility to implement cognition solely based on MRs and GPGICs—neural activity serving only the supportive role to execute simple sensory-motor loops.

2.1. Robust rerouting

Connectionism offers multiple possible mechanisms for temporarily rerouting the passage of activity through a neural network other than MRs and GPGICs (22,4,23,24). However, the present argument is that routing based on MRs and GPGICs is the only one powerful and flexible enough to fully implement cognition. GPGICs are able to maintain a new state of the network reliably and for a longer time: Once a GPGIC is activated it remains in the new state uninterrupted until it spontaneously returns back. This can last between several 100s of milliseconds up to several minutes (14). During this period, it is not possible by a voltage input to return the state of GPGIC back. Thus, the change made to routing cannot be disrupted or washed away. In contrast, other routing mechanisms, such as the phase of gamma oscillations (4) or recursive activity (25) are inherently unstable: A change in routing requires one to put a connectionist network into a fragile, fine-tuned state of the network's dynamics (for example by recurrently looping information within the network or setting the phase of an oscillation; 25,26). Such a state can easily be disrupted, washed away and replaced by another state. Consequently, connectionist mechanisms may require excessive resources to counteract this shortcoming and just to hold a small amount of information in a form of a newly rerouted network (27) (the scaling problem of the demands on resources by connectionist systems is discussed in detail below). Thus, rerouting based on GPGICs has the advantage of robustness and the amount of change that can be exerted onto the network. Therefore, GPGICs do not make minor routing adjustments to the flow of activity but they make a substantial, detailed selection of a subnetwork within the large physical network, and this selection remains in place for a certain period of time no matter what happens to the neural activity (Figure 2).

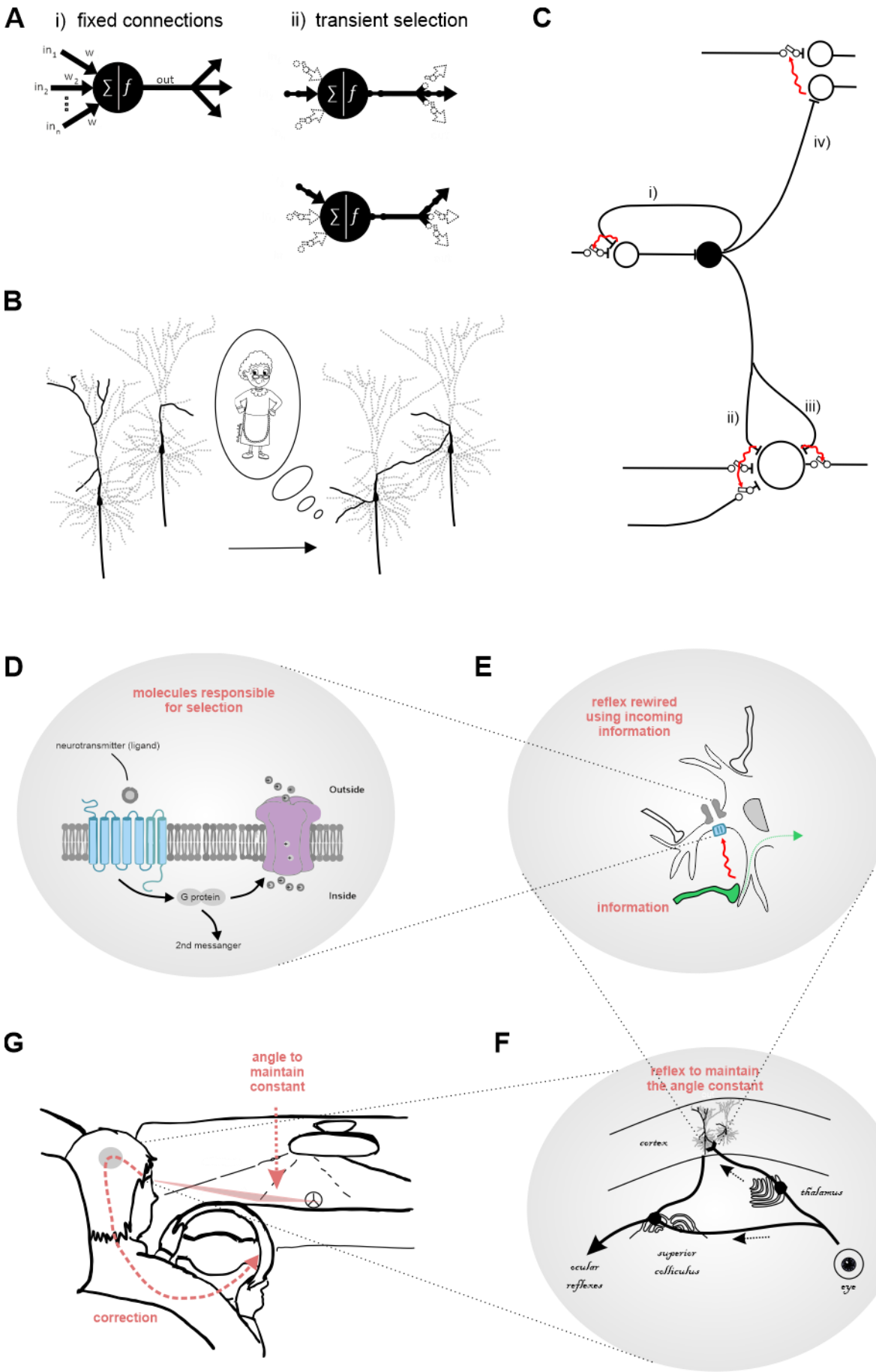

A
i) fixed connections
ii) transient selection
out
B
C
i)
ii)
iii)
iv)
D
molecules responsible for selection
neurotransmitter (ligand)
Outside
Inside
G protein
2nd messanger
E
reflex rewired using incoming information
information
G
angle to maintain constant
correction
F
reflex to maintain the angle constant
cortex
thalamus
superior colliculus
ocular reflexes
eye

Figure 1: **A putative ability of neural networks to transiently select subnetworks. A)** A schematic view on the transient selection hypothesis. i) The classical paradigm assumes that connection strengths are fixed across all tasks that a neuron performs. ii) Transient selection proposes that the routing of activity can temporarily change. Neurons are 'smart' enough to use their recent activity as a source of information for temporarily 'rewiring' their dendritic trees and axonal branches. **B)** A cognitive operation such as perception of a grandmother is executed by transiently altering the wiring of a network: To perceive a grandmother is to temporarily activate a subnetwork specific for interactions with grandmothers (a "grandmother network"). Black: wires opened for traffic; gray: wires closed for traffic. **C)** By sending outputs, a neuron not only inhibits or excites its targets but also provides information on the basis of which the target rewire itself. This information can be sent (i) backwards, (ii) laterally, (iii) forward and (iv) possibly also to neurons with which no direct synaptic connection is formed (disjoined interaction). **D)** Metabotropic receptors (MRs) and G protein-gated ion channels (GPGICs) are the membrane proteins proposed to be responsible for the transient selection capabilities of neural networks. An MR detects presence of a ligand in the extracellular space and releases a G protein which the activates a GPGIC; GPGIC opens or closes and by doing so alters the electrical properties of the membrane. **E)** One source of ligands are 'runaway' neurotransmitters resulting from nearby synaptic activity. **F)** As subnetworks are transiently selected, formed are novel temporary sensory-motor loops specific for the current task. **G)** Example of a temporary sensory-motor loop needed to drive a car – to keep the car within a lane by keeping constant a given visual angle making corrections by turning the steering wheel.

This robust transient selection allows for a further advantage: GPGICs are able to effectively accumulate changes in wiring as inputs arrive sequentially. While connectionist circuits are at the risk of losing information about the previous input in the sequence every time a new input arrives, this is not the case for GPGICs, as they have a select-and-forget property as far as neural activity is concerned. This means that new inputs can make new changes to wiring without the preceding changes being affected. Only such accumulation of changes in a lossless manner, without changes interfering one with another, can lead to a full implementation of cognition solely within MRs and GPGICs (Figure 2A).

Consequently, a network equipped with MRs and GPGICs can more effectively process information in a serial manner, than can connectionist-based recurrent systems (25,26). Serial processing is not only necessary for inputs of exclusively sequential nature such as language but also helps us understand how our brain is able to processes visual information also in a serial manner (28).

This also requires to re-think how different brain areas cooperate. Classically, it is assumed that different brain areas form a 'pipeline' hierarchy—each brain area corresponding to a processing stage (like from the brain area V1 to V2 … to IT cortex; Figure 2B, left). However, due to the sequential accumulation property of MRs and GPGICs, the internal processing in the brain cannot be organized as such an internal serial pipeline of processing stages. The reason is that the local changes made by MRs and GPGICs cannot be passed onto the next brain area. Instead, the transitive selection of pathways has effect only on the next round of the sensory input. Therefore, the new paradigm forces us to view the brain as an ultimate parallel device that achieves sequential processing by iteratively interacting with the environment. This is known as situated (embedded) (29) and enacted (30) cognition.

This also means that the passage of neural activity through the brain cannot conduct any elaborated forms of computations. Neural activity only executes simple sensory-motor loops from Figure 1B (Figure 2B, right). All of the brain areas contribute to these simple loops simultaneously – in parallel. Hence, the new paradigm emphasizes different set of anatomical connections as primary. While the connectionist paradigm emphasized cortical-cortical connections as the primary ones, the present paradigm emphasizes the connections feeding from subcortical areas to the cortex and back, considering the cortical-cortical projections only of secondary importance (Figure 2B, right).

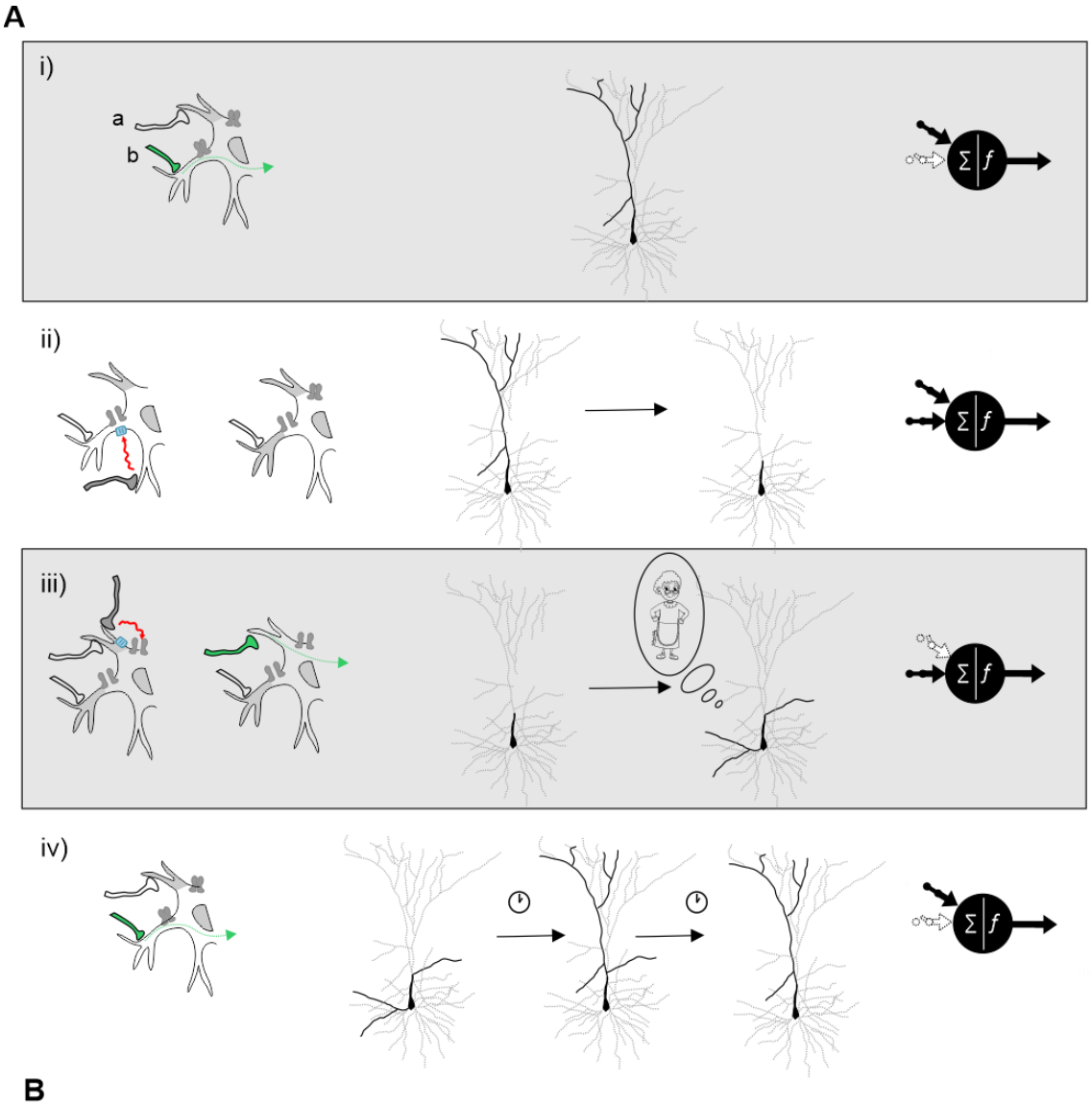

A
i)
a
b
ii)
iii)
iv)
B
Classical paradigm
output
area 3
area 2
area 1
input
Transient selection
area 3
area 2
area 1
input
output
External world

Figure 2. Sequential accumulation of changes to wiring. A) Changes serially accumulate and are eventually released. i) A default state of neurons: By default, some wires do not pass activity (*a*), while others do (*b*). ii) The actions of MRs and GPGICs cause closing of activity passage (or make it more difficult to pass) on one branch of the dendritic tree. iii) At another branch of the dendritic tree the actions of MRs and GPGICs enable passage through a previously closed part. The changes are being accumulated over events in ii) and iii), the neuron currently having considerably altered wiring properties. These changes may be a part of activating the 'grandmother' network. iv) Eventually, after some time, GPGICs return back to their initial default states. This return occurs without external inputs. **B)** Left: The sequential computation assumed by the classical paradigm: The computation results from one brain area are passed to the next brain area for further processing. Right: The results of transient rewiring cannot be passed to the next brain area; hence, the paradigm requires considering parallel operations of across the brain areas, sequential operations being achieved by closing a feedback cycle passing through the external world.

Consequently, the present paradigm is 'radical' in the sense that it does not move only some aspects of cognition to MRs and GPGICs, like it has been done in the past for short-term memory (e.g., 31), leaving the rest of cognition to the network activity. Instead, the present paradigm invites us to move the *entire* cognition to MRs and GPGICs. All of the various aspects of cognition ranging from attentional filtering to thinking, decision making and insights are proposed to emerge from the functions of MRs and GPGICs.

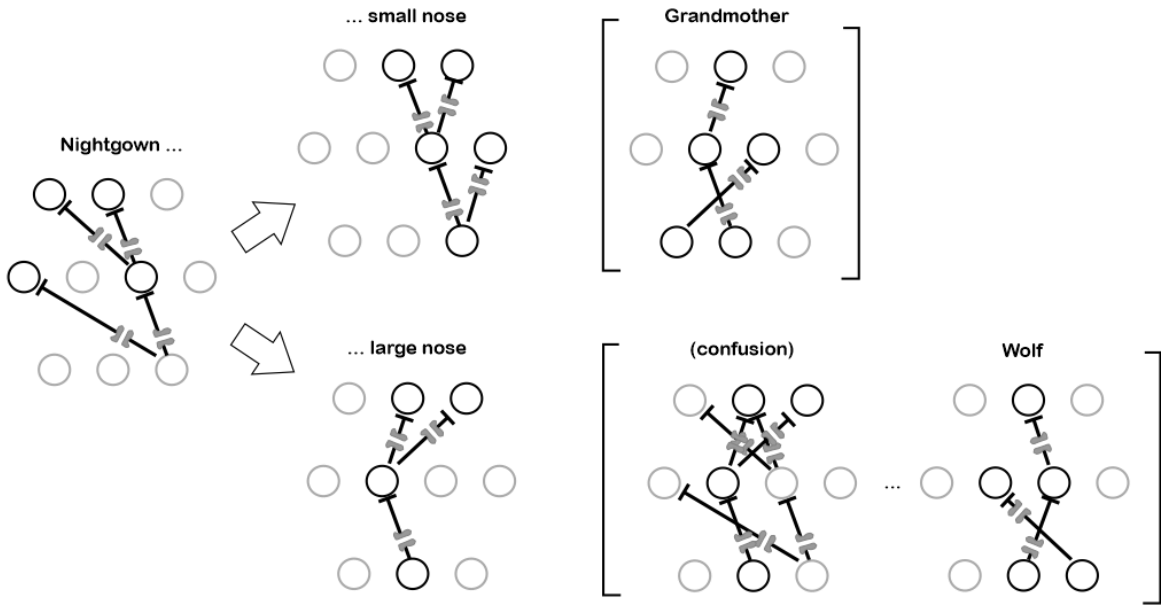


Figure 3. Schematic illustration of a temporal dynamics of transient selections. Cognition emerges from sequential accumulation of selections: Detecting nightgown followed by a small nose settles quickly into a familiar grandmother network; however, nightgown followed by a large nose creates an unsettling combination that demands a more extensive search for the correct familiar network, eventually resulting in the perception of a wolf.

A canonical scenario of fully engaging all aspects of our cognition involving a grandmother may look like this: (i) The earliest changes to wiring are made to the sensory systems to select potentially interesting inputs (this is attention); (ii) These changes direct subsequent sensory inputs which then detect in the sensory inputs a known object: a grandmother. The additional selection prepares the network for interacting with this familiar class of objects (this is perception); (iii) Further selection is needed to choose from the range of possibilities on how to interact, a particular one that will be used; should the grandmother be greeted, or

avoided? This may require several iterations of selection (here MRs and GPGICs are involved in thinking and problem solving); (iv) In some cases, the subnetwork that has been so far created by gradually accumulating changes may suddenly receive a major reorganization. A newly detected relation may be discovered: It is not a grandmother, it is a wolf dressed into grandmother's nightgown (this is an insight) (Figure 3); (v) Finally, the network may be selected in a way that causes vigorous actions: screaming, running away. In this case MRs and GPGICs select the circuits to boost the energy pumped into the organism and needed to accomplish the task of avoiding danger (this involves motivation, emotional reaction and decision making).

3. The problem of scaling intelligence

The true power of MRs and GPGICs as the explanation of our minds, is visible when considering the problem of scaling intelligence. The present argument states that animal-level and a human-level intelligence can only be reached by networks employing MRs and GPGICs and this is something impossible for the classical connectionist networks because these networks do not scale intelligence satisfactorily. Recent studies have elucidated the scaling limitations of connectionist networks: Investigations of the best performing connectionist networks, known as deep learning systems, revealed that advances in their intelligence can be only achieved if accompanied with an excessive increase in resources. Across all studies, these demands have been found to grow with a power law (32,33,27). To double the number of objects recognized or to reduce the error of classification by half, an equation of a form $r = i^a$ applies to describe the needed increase in the model size or the size of the training data set (r is the resources needed, i is intelligence and a is the exponent). The value of exponent a is always well above 1 (Figure 4A). One study estimated the exponent a to be ~9 for computer vision (33); this implies that doubling the intelligence of the state-of-the-art deep learning networks for vision (e.g., doubling the numbers of objects recognized without sacrificing accuracy), requires some 2^9 = ~500-fold increase in resources. Another study, implied an exponent of ~13 for language models (~8000-fold increase in resources for doubling the intelligence; Figure 4A) (27).

A question is whether the connections of the brain, which are different from those in deep learning models, can solve the scaling problem. This is unlikely. First, connectionist models of the brain have been around for a while (e.g., 2,22,34); if they were performing better than deep learning, this would have been noticed and would have been already used for AI technology. Second, computational neuroscience has never been able to show theoretical models that deal with anything but toy problems: To the best of the author's knowledge, no theoretical model has ever been able to cope with a real-life task. This suggests a generally poor scaling properties of connectionist systems. In addition, in Supplementary Materials there is an argument based on generalized logical XOR functions explaining why connectionism cannot scale effectively (applicable to both deep learning or biological networks).

These conclusions stand in contrast to the excellent scaling properties of biological brains. When comparing the brain sizes of different species, it is evident that only an increase in the brain size by a factor of 200 is sufficient to rise the intelligence from the level of a mouse to that of a human. Similarly, the factor of 4 distinguishes the sizes of the chimpanzee brain and of a human. Deep learning models cannot do much with such small increases in the number of parameters, as they would allow for an increase in intelligence at most by 80% (for a 200-fold increase) and by 20% (4-fold increase). If one needs to increase the intelligence of deep learning systems by for example a factor of 1000, which would be a more realistic expectation when comparing a human to a mouse brain, this would require at

least a $1000^a = 10^{27}$ if not a 10^{39}-fold increase in the brain size, which would require a human brain of the size of a celestial body.

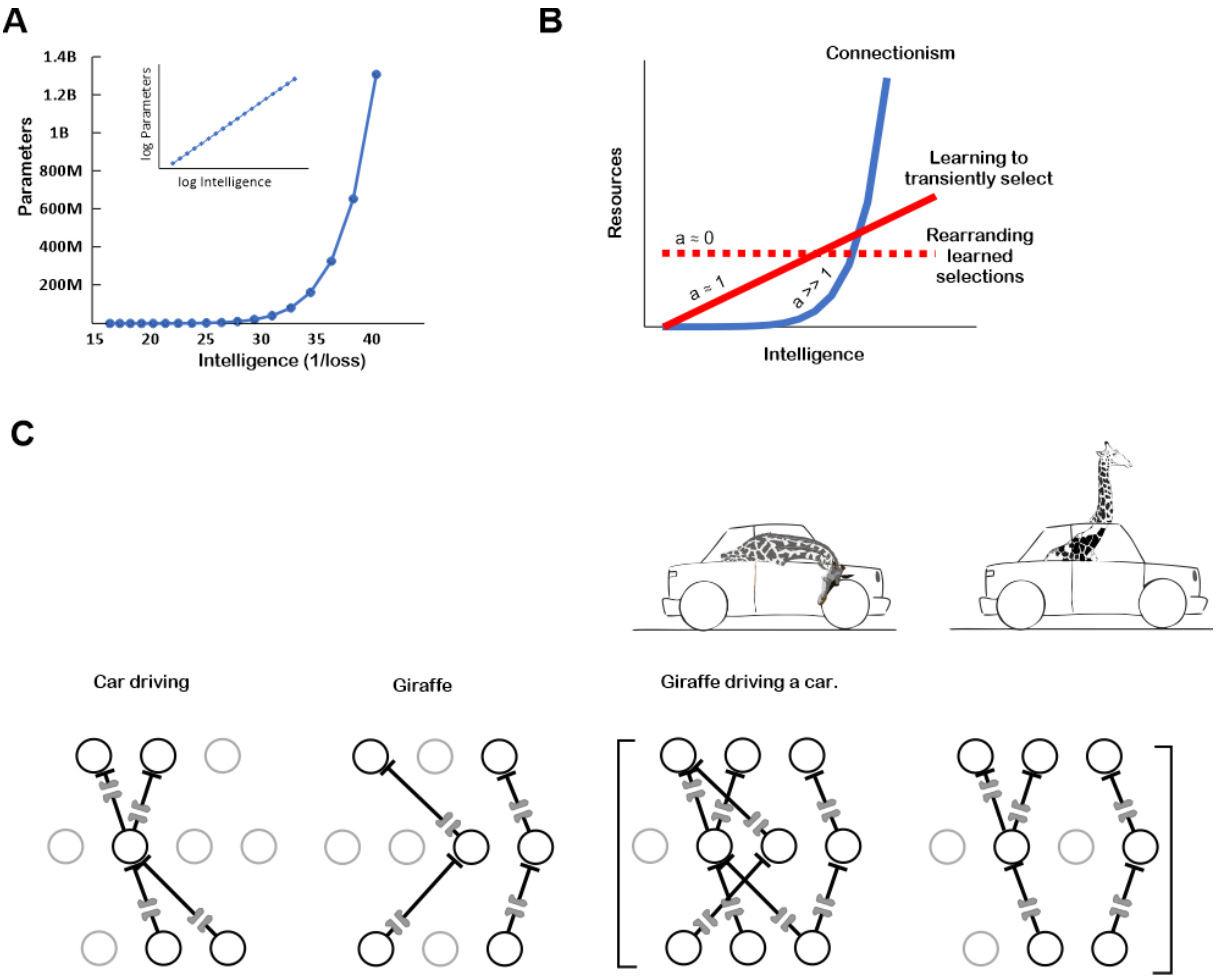


Figure 4. Scaling intelligence in connectionist systems vs. transient selection. A) A power law relation between the intelligence of a transformer model for language processing (intelligence defined as 1/loss) and the number of parameters in the model. Adapted from (27). Inset: the same graph shown in a log-log plot, the straight line indicating a power law. **B)** Demands on resources for various paradigms. Blue: In connectionism the demands on resources grow with power law where the exponent *a* is much larger than one. With transient selection, the resources grow linearly for learning (solid red line) and remain unchanged for recombining the learned pieces of knowledge ($a = 0$; dashed red line). **C)** Schematic illustration of a creative composition of existing pieces of knowledge. If one is presented with the idea of 'Giraffe driving a car' for the first time, the subnetworks for 'car driving' and for 'giraffe' interact. Finding a solution may take multiple iteration. The first subnetwork may not be optimal (e.g., inconvenient for the giraffe). In the next iteration the knowledge about cars having sunroofs becomes relevant, allow to find a better way for a giraffe to drive a car.

Even more excessive demands can be obtained when comparing humans and machines. The number of visual objects a state-of-the-art connectionist machine may distinguish reliably is in the order of 10^3. In contrast, biological brains have the capability to understand and accurately perceive visual objects created by novel arrangements of elements forming constellations never seen before. An assessment of the total variety of situations that can be

correctly perceived by an adult human lead to a lower-bound estimate of 10^{48} and the upper bound of 10^{64}, all without having to resort to new learning (35). By extrapolating the power laws ruling connectionist machines, the discrepancies remain so vast (in the order of 10^{45} to 10^{61}) that to catch up with the human brain, state-of-the-art deep learning systems would need to exceed the size of the known universe. If biological neural networks were limited to the connectionist paradigm i.e., without MRs and GPGICs, would likely scale even worse.

Thus, it is a must for biological systems to scale their intelligence better than what connectionism can offer. Otherwise, they could not survive in real-world environments, which are complex and often hostile. Biological systems exhibit two intelligence-scaling properties that respectively have power law exponents 1 and 0. We acquire new knowledge i.e., learn, with the exponent of about 1 ($a \approx 1$): The number of facts or skills learned is about a linear function of the time we spent learning them; adding to our repertoire a new fact or a skill takes about as much effort as it took to learn previous fact or skill. This is not the case with connectionist systems which require exploding amounts of training examples and of the training time ($a >> 1$). This is largely due to the need to retrain together new and old items (36) because otherwise, newly learned items start erasing previously learned items, leading to catastrophic interference (37). Biological brains do not have this problem and the present argument is that this is due to cognition being implemented by MRs and GPGICs.

The second intelligence-scaling property of biological systems is the near-zero exponent resulting from the ability to rearrange previously acquired pieces of knowledge into a new combination. Many problems that would be considered new for connectionist networks are in fact just variations of previous problems and can be solved by reusing components of previously learned skills. If an intelligent system is able to disentangle any given situation into the correct components, like juggling mental Lego pieces, and is able to learn these Lego pieces instead of learning a solution to the problem by rote, then a possibility emerges to reuse these Lego pieces to solve new problems by rearranging them into novel combinations. This enables humans to drive cars under varying driving conditions without having to practice millions of examples of such driving conditions – in contrast to the connectionist systems. This also enables animals navigate and survive the varying conditions in complex and hostile environments. This Lego-like decomposition of skills also makes our language productive (38): We can quickly describe new situations and a listener can quickly understand them, even if these situations have never occurred before: "*an astronaut using its oxygen tank to pump up the tires of the space bicycle*" or "*a dinosaur needs privacy when using a dinosaur toilet*". There is a vast number of possible combinations of Lego elements producing situations that we effortlessly understand; the lower bound of this number has been estimated to 10^{48} (35). These novel arrangements are for our brains so easy and often take so little time that the whole effort is negligible in comparison to the efforts connectionists systems need to undertake to learn to deal with such new situations. Biological cognition with its attention, perception, working memory and so on is the process of rearranging these Lego pieces; these cognitive operations effectively reduce the power law exponent of to $a \approx 0$ (Figure 4B).

The present claim is that both reductions in the power law exponent, the $a \approx 1$ for learning and $a \approx 0$ for cognition are achieved through cognition relying on MRs and GPGICs i.e., through cognition founded in the capability to transiently and robustly select subnetworks. The mentioned Lego pieces are manifested in the selected subnetworks. The $a \approx 1$ is proposed to be achieved by applying plastic changes only to the selected pathways; ensuring that the majority of pathways, those unselected, do not get affected by the learning process; this should prevent catastrophic interreference and make it possible to learn from a small number of examples. What is learned is a new subnetwork. This network takes at first several iterations of accumulation to get the subnetwork right; after learning, the same

selection is made a lot faster, by fewer iterations of by a single cue (the underlying learning rules are currently not known and still need to be researched (39)). That way, also a new Lego piece is learned which can be used to expand the domain within which the cognition operates, boosting the intelligence of the system as whole even further: now $a \approx 0$ within a one-step broader domain. This is how a person's or animal's intelligence grows iteratively through their lifetime: the existing Lego pieces enable effective learning of new Lego pieces.

This iterative nature of learning where previous knowledge assists acquisition of new knowledge, explains why we, unlike connectionist systems, need to learn the basic concepts before learning more advanced skills or knowledge items: counting and addition is learned before calculus, walking before running or playing soccer, and so on. Finally, the Lego pieces that can be flexibly combined are the foundations of our semantics and of our ability to acquire new concepts. Figure 4C illustrates how $a \approx 0$ leads to our ability to solve problems by insights.

4. Observing transient selection in electrophysiological experiments

The argument for transient selection of networks would be incomplete if there was no data supporting the proposal. Do we observe in the brain activity responses that can be interpreted as a signature of a transient selection of subnetworks? The very first observation once an electrode is placed within a nervous system is that neural responses to stimuli are not constant to constant stimuli. Instead, the responses tend to gradually secede after initial vigorous activity (40,41). This phenomenon can be referred to as fast adaptation (41) or transient responses (42) and may well be caused by activation of GPGICs and a consequent reduction of the ability of neuron membranes to conduct signals. A related common observation is that the response are never identical even if identical stimuli are presented. Instead, neural responses are marked by high variability (43,44). This suggests that the underlying network continually changes its properties, which may well be induced by the operations of MRs and GPGICs.

Another relevant phenomenon is the finding that neural activity following the removal of a stimulus sometimes contains more stimulus-related information (i.e., more stimulus specificity) than the activity during the stimulus presentation. In our studies we found that the information about the identity of flashed stimulus was preserved for at least 500 milliseconds after the stimulus flash was over (45,46) (Figure 5A). This information may not simply reverberate within a classical connectionist networks (as we originally suggested) but may involve transient changes in the reverberating pathways selected by the operations of MRs and GPGICs, prolonging thereby the duration of the memory effect.

Spontaneous activity in intact brains, slices and cultures of neurons (47,48,49,50) is yet another phenomenon that can be explained by transient rewiring of these subnetworks. Spontaneous activity has properties of avalanches (51,52) and the 'fuel' as well as a trigger for an avalanche may be changes in the membrane properties induced by GPGICs returning back to their deactivated state, lowering the activation thresholds. A triggered avalanche of activity may then activate a set of other GPGICs, which after some time return back to deactivated state and so on – the pattern of activity irregularly repeating.

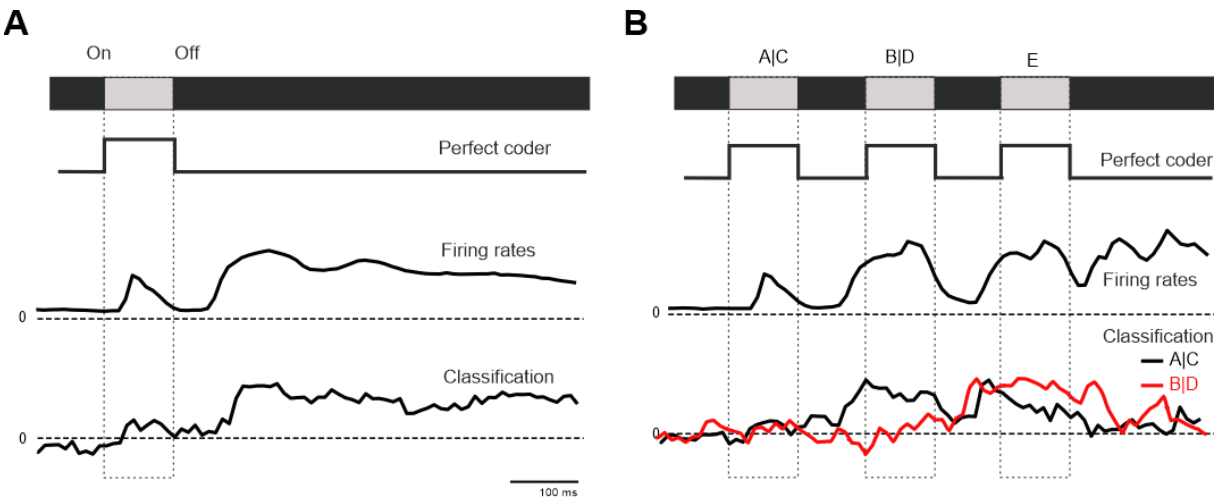


Figure 5. **Visual cortex maintains information about recent stimuli**. **A)** Responses to a flashed stimulus by a theoretically perfect coder compared to the response of a population of neurons to such a stimulus in the primary visual cortex (Firing rates), and the stimulus-related information contained within these responses (Classification). Information about the identity of a stimulus is available for several hundreds of milliseconds after the stimulus has been removed. **B)** The same as A) but for a sequence of flashed stimuli. Responses to any stimulus in a sequence contain information about the recently presented stimuli (for example, responses to B|D containing information about A|C). Adapted from 45.

Finally, in one study, we have even observed responses that may present a signature of the process of sequential accumulation of transient selection of pathways (45). We flashed sequences of three stimuli and the responses to any given stimulus contained not only information about that stimulus but also about the immediately preceding stimulus (Figure 5B). This indicates superposition of information which was accumulated over time. This phenomenon is difficult to explain simply by activity reverberating within a connectionist system. In fact, to the best of author's knowledge, no connectionist model has been made to date capable of accumulating information in such a manner. Transient selection of pathways based on MRs and GPGICs elegantly explains these results.

5. Conclusions

Transient selection is a different approach to explaining the mind from the classical connectionist-based approach. While the classical paradigm relies on network connections for storing information and on membrane voltages as the computation medium, the new paradigm relies on MRs for detecting the conditions for executing the 'computation' and on the states of GPGICs for holding the result of that 'computation'. The mental state is defined by the states of GPGICs. The state of voltages does not matter so much, as it can be discarded.

Connectionism considers the process of setting the membrane voltages as the 'algorithm' that executes the thought. Transitive selection, in contrast, considers this aspect of the nervous system only as a shadow of another process that operates behind the voltages and that possesses the true power of intelligence: This hidden intelligent entity sets up the pathways for the voltage-based computations. Counterintuitively, this intelligent entity is not a compact machine that can be described as a closed box with its inputs and outputs.

Instead, the intelligent entity constitutes of a set of small elements that are distributed across the entire nervous system and that, on the surface, seem to operate independently. Nevertheless, they manage to coordinate and by doing so form a mind.

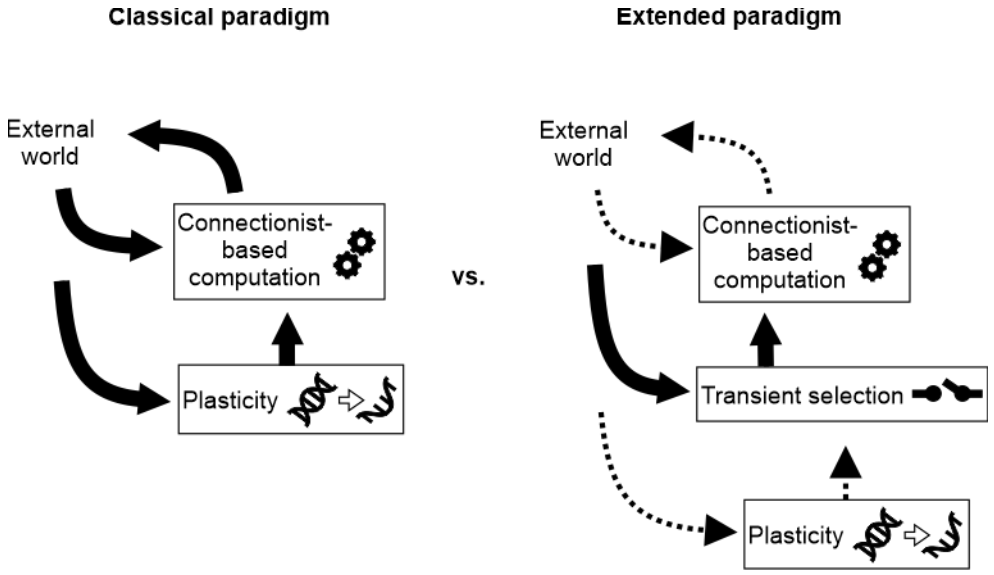


Figure 6. **Transient selection paradigm extends the classical paradigm with an additional group of mechanisms.** Left: Two major sets of adaptive mechanisms as presumed by the connectionist paradigm: one set is for computations based on network activity and the other set for plastic changes to that network. Both mechanisms receive feedback from the external world. Right: Transient selection introduces a third set of adaptive mechanisms, transient selection, operating in-between the plasticity and connectionist-based computations and also receiving feedback from the external world.

The mechanisms by which MRs and GPGICs can coordinate across the nervous system are addressed by the theory of practopoiesis (53). The theory starts by presenting an abstract, cybernetics-based, account of how neural plasticity manages to coordinate neural activity. These two sets of mechanisms are found to comply with the same set of abstract rules (53). One example rule is the need for both to receive feedback from the external world (Figure 6, left). By analyzing those rules, the theory envisioned a third set of mechanisms operating in-between i.e., sitting in a sandwich between plasticity on one side and connectionist computations on the other side (Figure 6, right). This third set of mechanisms, would be coordinated by plasticity but it would itself act as a coordinator of the neural activity. The mechanism was originally named *anapoiesis*, and has been presently identified as MRs and GPGICs. Practopoiesis offers a number of interesting insights and should be useful for future research on MRs and GPGICs, including the investigations of the underlying learning mechanisms (39).

The new paradigm based on transitive selection of pathways opens a new approach towards the problem of explaining natural intelligence. The paradigm also offers implications for development of artificial intelligence (AI) technology that would be similar to biological intelligence.

Acknowledgments
The author would like to thank Gerald Hahn and Visvanathan Ramesh for feedback on an earlier version of the manuscript. **Conflict of interest:** The author is working on a patent for an AI technology for transient selection of pathways within artificial neural networks.

References

1. Lloyd, D., Consciousness: A connectionist manifesto. *Minds and Machines* **5(2)**, 161–185 (1995).

2. Grossberg, S. (2013). Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. Neural networks, 37, 1-47.

3. Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. Neural computation, 29(1), 1-49.

4. Fries, P., Nikolić, D., & Singer, W. (2007). The gamma cycle. Trends in neurosciences, 30(7), 309-316.

5. Massaro:, D., Some criticisms of connectionist models of human performance,. *Journal of Memory and Language* **27** (2), 213-234 (1988).

6. Van Hemmen, J. L. . &. S. T. J. (.)., *23 problems in systems neuroscience* (Oxford University Press., 2005).

7. Fodor, J. A. & Pylyshyn, Z. W., Connectionism and cognitive architecture: A critical analysis. *Cognition* **28** (1-2), 3-71 (1988).

8. Kumar, A. . V. I. . A. A. . &. B. C., Challenges of understanding brain function by selective modulation of neuronal subpopulations. *Trends in Neurosciences* **36** (10), 579-586. (2013).

9. Block, N. . & Stalnaker, R., Conceptual analysis, dualism, and the explanatory gap.. *The Philosophical Review* **108** (1), 1-46 (1999).

10. Baars, B. J., Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in brain research* **150**, 45-53 (2005).

11. Block, N., On a confusion about a function of consciousness. *Behavioral and brain sciences* **18** (2), 227-247 (1995).

12. VanRullen, R. . &. K. R., Deep learning and the global workspace theory. *Trends in Neurosciences* **44** (9), 692-704 (2021).

13. Burke Jr, K. J., & Bender, K. J. (2019). Modulation of ion channels in the axon: mechanisms and function. Frontiers in cellular neuroscience, 13, 221.

14. Inanobe, A., & Kurachi, Y. (2014). Membrane channels as integrators of G-protein-mediated signaling. Biochimica Et Biophysica Acta (BBA)-Biomembranes, 1838(2), 521-531.

15. Breitwieser, G. E. (1991). G protein-mediated ion channel activation. Hypertension, 17(5), 684-692.

16. Mark, M. D., & Herlitze, S. (2000). G-protein mediated gating of inward-rectifier K+ channels. European Journal of Biochemistry, 267(19), 5830-5836.

17. Lüscher, C., Jan, L. Y., Stoffel, M., Malenka, R. C., & Nicoll, R. A. (1997). G protein-coupled inwardly rectifying K+ channels (GIRKs) mediate postsynaptic but not presynaptic transmitter actions in hippocampal neurons. Neuron, 19(3), 687-695.

18. Basith, S., Cui, M., Macalino, S. J., Park, J., Clavio, N. A., Kang, S., & Choi, S. (2018). Exploring G protein-coupled receptors (GPCRs) ligand space via cheminformatics approaches: impact on rational drug design. Frontiers in pharmacology, 9, 128.

19. Koyrakh, L., Luján, R., Colón, J., Karschin, C., Kurachi, Y., Karschin, A., & Wickman, K. (2005). Molecular and cellular diversity of neuronal G-protein-gated potassium channels. Journal of Neuroscience, 25(49), 11468-11478.

20. Chen, X., & Johnston, D. (2005). Constitutively active G-protein-gated inwardly rectifying K+ channels in dendrites of hippocampal CA1 pyramidal neurons. Journal of Neuroscience, 25(15), 3787-3792.

21. LoRusso, E., Hickman, J. J., & Guo, X. (2019). Ion channel dysfunction and altered motoneuron excitability in ALS. Neurological disorders & epilepsy journal, 3(2).

22. *Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1988). Parallel distributed processing (Vol. 1, pp. 354-362). New York: IEEE.*

23. Friston, K. J. (2011). Functional and effective connectivity: a review. Brain Connect. 1, 13–36. doi: 10.1089/brain.2011.0008.

24. Lisman, J. E., Fellous, J. M., & Wang, X. J. (1998). A role for NMDA-receptor channels in working memory. Nature neuroscience, 1(4), 273-275.

25. Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. Neural computation, 14(11), 2531-2560.

26. Donner, R. V., Zou, Y., Donges, J. F., Marwan, N., & Kurths, J. (2010). Ambiguities in recurrence-based complex network representations of time series. Physical Review E, 81(1), 015101.

27. Kaplan, J. . M. S. . H. T. . B. T. B. . C. B. . C. R. .. &. A. D., 2020.

28. Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive psychology, 12(1), 97-136.

29. Gibson, J. J. (1977). The theory of affordances. Hilldale, USA, 1(2), 67-82.

30. Maturana, H. R., & Varela, F. J. (1987). The tree of knowledge: The biological roots of human understanding. New Science Library/Shambhala Publications.

31. Goelet, P., Castellucci, V. F., Schacher, S., & Kandel, E. R. (1986). The long and the short of long–term memory—a molecular framework. Nature, 322(6078), 419-422.

32. Meir, Y. . S. S. . H. S. . K. K. . B.-N. I. . G. A. . &. K. I., Power-law scaling to assist with key challenges in artificial intelligence.. *Scientific reports* **10** (1), 1-7. (2020).

33. Thompson, N. C. . G. K. . L. K. . &. M. G. F., 2020.

34. Moca, V. V., Nikolić, D., Singer, W., & Mureşan, R. C. (2014). Membrane resonance enables stable and robust gamma oscillations. Cerebral cortex, 24(1), 119-142.

35. Nikolić, D., Why deep neural nets cannot ever match biological intelligence and what to do about it?. *International Journal of Automation and Computing* **14** (5), 532-541 (2017).

36. Zeng, G., Chen, Y., Cui, B., & Yu, S. (2019). Continual learning of context-dependent processing in neural networks. Nature Machine Intelligence, 1(8), 364-372.

37. McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In Psychology of learning and motivation (Vol. 24, pp. 109-165). Academic Press.

38. Yang, C., Crain, S., Berwick, R. C., Chomsky, N., & Bolhuis, J. J. (2017). The growth of language: Universal Grammar, experience, and principles of computation. Neuroscience & Biobehavioral Reviews, 81, 103-119.

39. Nikolić, D., in *Closed Loop Neuroscience* (2016), pp. 53-65.

40. Benda, J. (2021). Neural adaptation. Current Biology, 31(3), R110-R116.

41. Dean, I., Robinson, B. L., Harper, N. S., & McAlpine, D. (2008). Rapid neural adaptation to sound level statistics. Journal of Neuroscience, 28(25), 6430-6438.

42. Sun, W., & Dan, Y. (2009). Layer-specific network oscillation and spatiotemporal receptive field in the visual cortex. Proceedings of the National Academy of Sciences, 106(42), 17986-17991.

43. Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S.,. & Shenoy, K. V. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nature neuroscience, 13(3), 369-378.

44. Dinstein, I., Heeger, D. J., & Behrmann, M. (2015). Neural variability: friend or foe? Trends in cognitive sciences, 19(6), 322-328.

45. Nikolić, D. . H. S. . S. W. . &. M. W., Distributed fading memory for stimulus properties in the primary visual cortex. *PLoS Biology* **7** (12), e1000260. (2009).

46. Lazar, A. . L. C. . F. P. . S. W. . &. N. D., Visual exposure enhances stimulus encoding and persistence in primary cortex. *Proceedings of the National Academy of Sciences* **118** (43) (2021).

47. Mao, B. Q., Hamzei-Sichani, F., Aronov, D., Froemke, R. C., & Yuste, R. (2001). Dynamics of spontaneous activity in neocortical slices. Neuron, 32(5), 883-898.

48. *Creutzfeldt, Otto Detlev, and Otto Creutzfeldt. Cortex cerebri: performance, structural, and functional organization of the cortex. Oxford University Press, USA, 1995.*

49. Tsodyks, M., Kenet, T., Grinvald, A., & Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. Science, 286(5446), 1943-1946.

50. Mazzoni, A., Broccard, F. D., Garcia-Perez, E., Bonifazi, P., Ruaro, M. E., & Torre, V. (2007). On the dynamics of the spontaneous activity in neuronal networks. PloS one, 2(5), e439.

51. Hahn, G., Petermann, T., Havenith, M. N., Yu, S., Singer, W., Plenz, D., & Nikolić, D. (2010). Neuronal avalanches in spontaneous activity in vivo. Journal of neurophysiology, 104(6), 3312-3322.

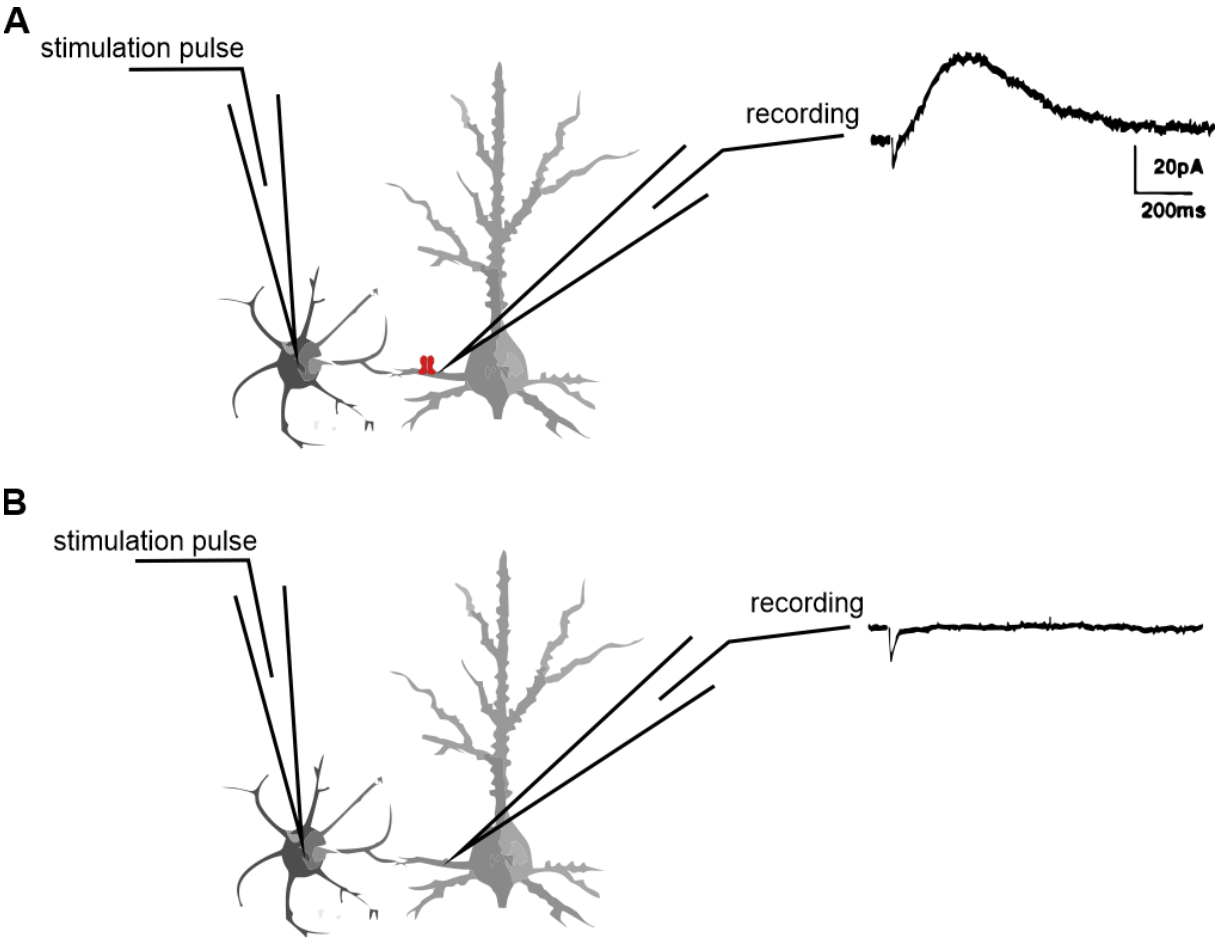
52. Priesemann, V., Wibral, M., Valderrama, M., Pröpper, R., Le Van Quyen, M., Geisel, T.,. & Munk, M. H. (2014). Spike avalanches in vivo suggest a driven, slightly subcritical brain state. Frontiers in systems neuroscience, 8, 108.

53. Nikolić, D., Practopoiesis: Or how life fosters a mind. *Journal of Theoretical Biology* **373**, 40-61. (2015).

Supplementary Materials for Nikolić, D. ”Where is the mind within the brain? Transient selection of subnetworks by metabotropic receptors and G protein-gated ion channels”

How GPGICs transiently rewire networks

An example of how the activity of GPGICs affects computations is shown in Supplementary Figure 1. If an inhibitory neuron is stimulated, two types of inhibitory currents are produced in a pyramidal cell of a normal healthy mouse. First there is a short lasting one and followed by a long lasting one (Supplementary Figure 1A). However, if the mouse is genetically altered and the gene for a specific type of GPGIC (the type is GIRK2 in this case) is knocked out, the responses of the pyramidal cell to a stimulation of an inhibitory cell show only the fast inhibitory currents while the slow ones are missing (Supplementary Figure 1B). This second, slow component can change the circuit properties for a certain period of time. The stimulus can that way affect how subsequent inputs are being processed. If the GPGIC is missing, inhibition is still present, but only short-lived, the effects being ‘forgotten’ rapidly.

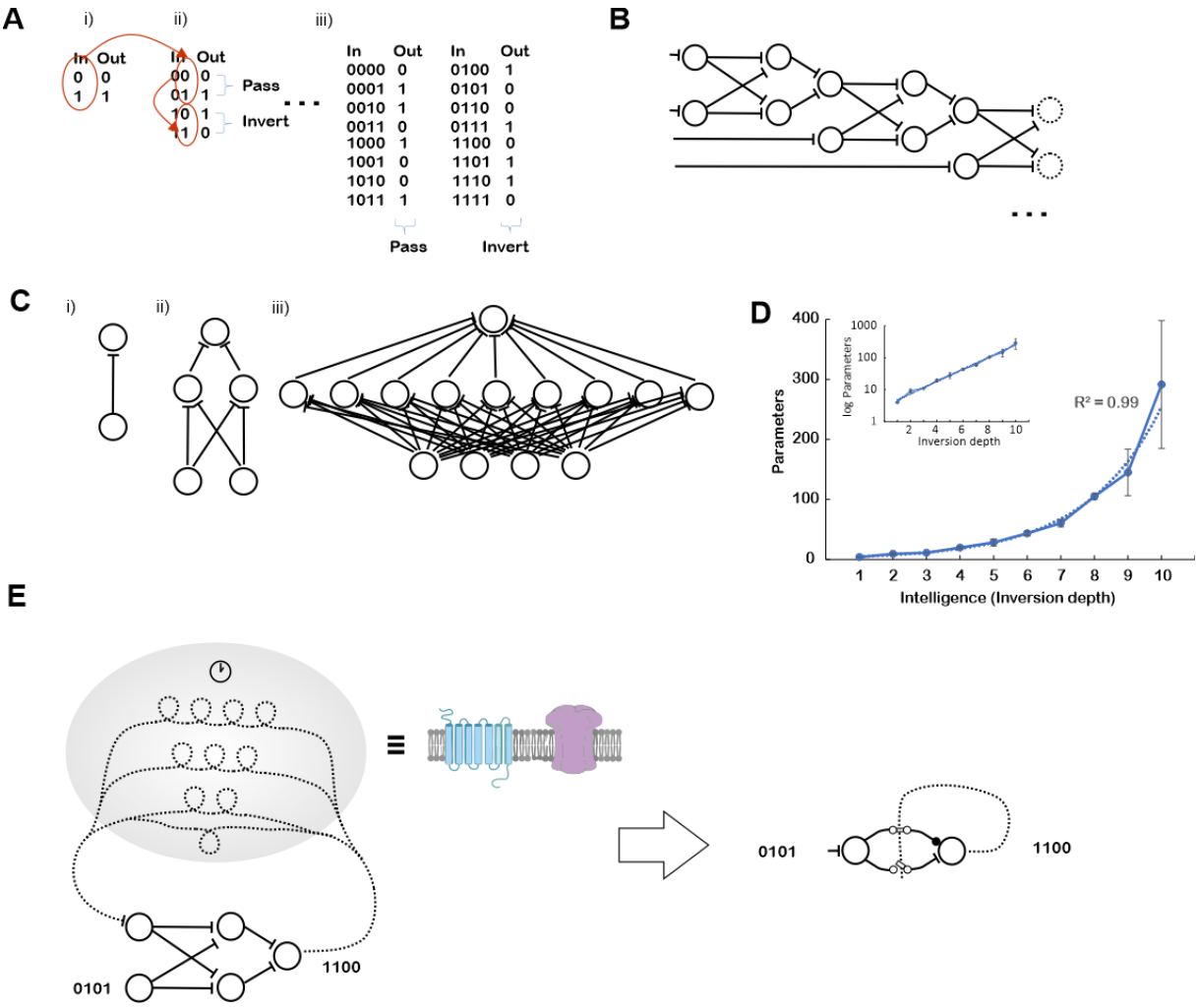


Supplementary Figure 1. Inhibitory postsynaptic currents measured form a CA1 pyramidal cell in hippocampus. Stimulation pulse is delivered to an inhibitory neuron. **A)** Responses in a cell of a normal mouse (wild-type). **B)** Responses of a knock-out mouse lacking GIRK2 protein (red only in A). Adapted from (17)

The XOR argument

A yet another perspective on how MRs and GPGICs empower neural networks can be provided by considering implementations of XOR operations. XOR operations are a part of the real life: "one may spend money on a nice dinner or on a theater play, but not both." or "animals either have four legs or two legs and a pair of wings, but not both", and can be generalized into a multi-bit version called inverters (Supplementary Figure 2A). It has been well established that connectionist systems are able to implement XOR functions. However, there are some limitations that prevent connectionist systems to scale their intelligence for any real-life problems that require XOR operations. First, theoretically, inverters i.e., complex XOR functions, can at best be implemented with linear scaling in the connectionist systems (Supplementary Figure 2B). This is a problem because this means that, even under theoretically most optimal conditions, connectionist systems cannot reduce the amount of computation to less than $a = 1$. With such scaling it is not possible to distinguish to the order of 10^{48} categories. The second limitation makes this problem even worse: the learning mechanisms are notoriously incompetent in finding the correct connection patterns needed for implementing XOR mapping functions; the practical scaling properties of connectionist systems are much worse than linear when it comes to XOR. Supplementary Figures 2C and D illustrate the results of a computational experiment in which the smallest possible network sizes were found still able to learn inverters of different depths by using a state-of-the-art learning algorithm. The required sizes of the networks exploded (see Supplementary materials for details of the experiment).

Supplementary Figure 2E illustrates how the inverter problem can be solved with $a = 0$ by extending the assumptions of connectionism: by adding delay lines at the right space the inverter problem can be solved with effective scaling. Infinite length inverter can be created. Such delay lines would be resource demanding if implement in the brain simply as long wires; a more effective approach to creating a local memory mechanism of short duration is to place a GPGIC onto the membrane and activate it through an MR (Supplementary Figure 2E, right; see Supplementary materials). Thus, one way of thinking about MRs and GPGICs is as local machinery for storing shot-lasting memories in a non-linear form, effectively applying XOR functions on sequential inputs. If arranged accordingly across the nervous systems (by currently unknown learning rules), this machinery conducts operations sufficient to implement cognition.



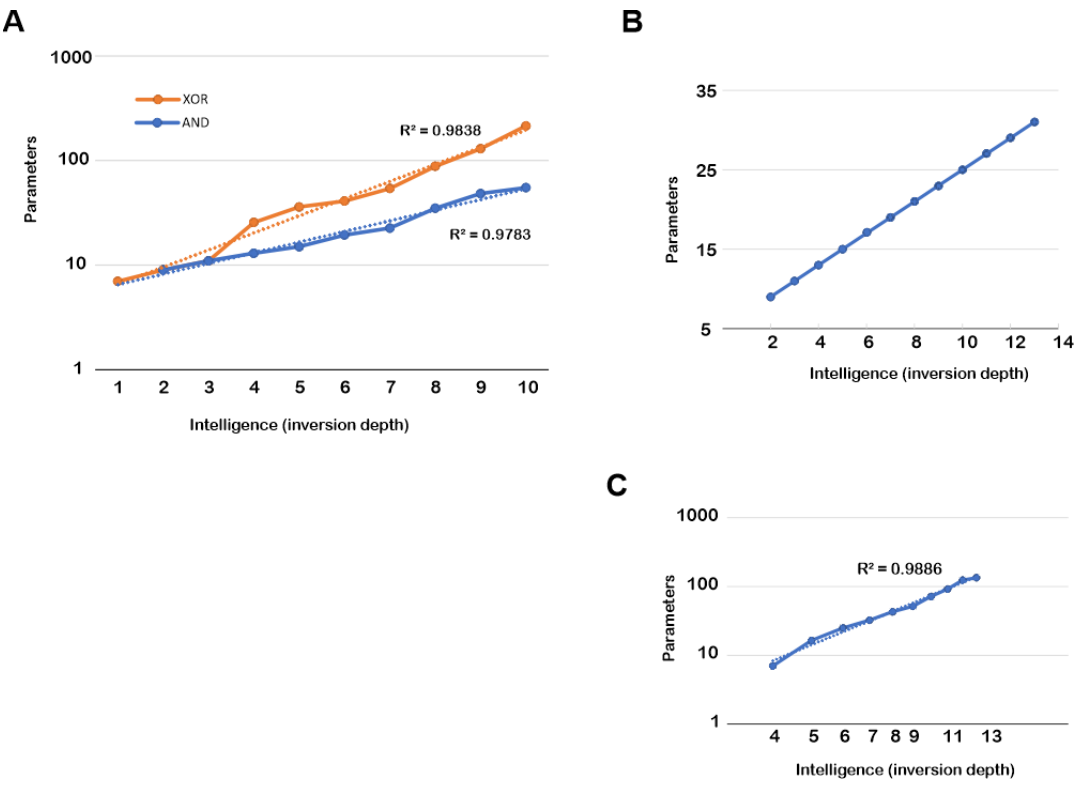
Supplementary Figure 2. Transient selections for XOR problems. A) The XOR problem generalized to larger number of bits, known as inverters. i) to iii): inversion depts of 1, 2 and 4 bits. **B)** In theory, connectionist networks can implement the generalized XOR problems (inverters) by linearly stacking XOR circuits. **C)** The results of an experiments: The sizes of networks discovered for inversion depths i) I to iii) in A. **D)** The number of parameters in models for experiments in C) for inversion depths up to 10 bits. Inset: the same data shown in a log-linear graph; the straight line indicating an exponential growth in the demands on resources. Vertical bars: standard deviation across 10 repetitions. **E)** Left: By combining delay lines with a recursive connection, a single XOR network can implement a generalized XOR of unlimited inversion depth. Delay lines can be implemented elegantly by MRs and GPGICs. Right: a simple recursive circuit equipped with gates for transient pathway selection produces generalized XOR function (more details in Supplementary materials).

Inverters are generalizations of the XOR function to an unlimited number of bits. With each additional bit an inverter can be generated by applying the following rule: If the bit is zero, the outputs will be exactly the opposite than if the bit is one. The function poses difficulties when implementing in a mapping function of a connectionist network because each XOR transformation is non-monotonic. Non-monotonic functions are a subset of non-linear functions and are more difficult to implement than monotonic non-linear functions.

While an inverter function can easily be constructed by within a connectionist network, learning of such a function proves difficult. The reason is that the non-monotonic relationships cannot directly be detected and implemented by the learning rules applied to connectionist networks. Instead, a connectionist network largely needs to stumble upon such an implementation i.e., by chance. The learning rule can only select and sharpen already existing inverter capabilities of networks. As a result, we found that the sizes of randomly initialized networks able to learn inverting functions grow exponentially as a function of the inversion dept (the number of bits in the function), as show in our study.

We trained network of different sizes and configurations to learn inverter functions of different depths. The depts of the inverter functions ranged from 2 to 10 bits. We used a step-wise increase function, whereby we increased the size of the hidden layer(s) b one neuron. A network would be initialized with random numbers and would be trained until it reached a learning criterion. We used two types of criteria: a low value for loss and accuracy, which needed to be perfect. The low loss criterion required the network to reach a pre-specified value for average loss, which was set to: 0.00001. The exact accuracy criterion required the network to categories the inputs into 1 or 0 output according to the inverter function correctly for all possibly combinations of inputs. The size of the network was increased stepwise; if a network of a given size did not converge, the size of all layers in the network increased by one neuron and the training procedure was repeated. The size of the network was limited to maximum of 150 neurons. To speed up the experiment, an early stop rule was implemented such that a training procedure was declared as not converging if loss did not improve a pre-specified number of steps (patience). Patience varied across experiments and ranged between 25 and 125. The maximum number of epochs allowed for training varied between 4,000 to 200,000. All layers were fully connected and the transfer function was sigmoid. The models had only one output neuron: If output < 0.5 the output was converted to 0, otherwise to 1.

The experiment was performed on GPU GTX 2070 SUPER with 8 GB or RAM and using PyTorch framework version 1.10.1. Learning rules was Adam with the learning rate 0.03. To assess the variability of the network sizes that converged (i.e., learned the inverter function) the procedure was repeated for each inversion depth until the network converged 10 times.



Supplementary Figure 3. **A)** The results to the XOR task for the criterion perfect accuracy compared to an AND task with the same criterion. A straight line in a log-linear plot indicates exponential growth in model size. **B)** The same as in B) but for learning logical OR function. A straight line in a linear-linear plot indicates a linear growth in needed resources. **C)** By mixing AND and OR functions, the best fit was obtained by a power law function, as indicated by a straight line in a log-log plot.

The exponential growth in resources reported in the main text was obtained by the low-loss criterion. This result was robust to a change in the criterion: We also obtained an exponential growth in needed resources when perfect accuracy was used as a criterion (Supplementary Figure 3A, orange). We also growth in resources of AND and OR logical operations. Logical AND operation also required an exponential growth in resources, but this growth was much smaller – i.e., it had a much smaller exponent (Supplementary Figure 3A, blue). Therefore, AND is a lot easier operations to learn but it still grows exponentially.

These results implied that AND and XOR logical operations could not explain the power law growth in resources in real-life situations (vision and language). Exponential growth is much more 'aggressive' than power law; exponential growth explodes faster. (The difference between the two can be also noted by how the two can be converted into straight lines in plots: to present an exponential function as a straight line, the scale on the ordinate needs to be converted to the opposite of the exponent i.e., to the logarithm (as in Supplementary Figure 3A). In contrast, to visualize a power law function as a straight line, the cases of both abscissa and ordinate need to be presented as a logarithm.) To account for real life's power law growth, we looked at the OR logical operation which turned out to present linear demands on resources (Supplementary Figure 3B). That is, an OR function is the easiest one to learn by deep learning (connectionist) networks. We then propose that the power law function in real life results from a mixture of OR, AND and XOR operations that occur in real

life situations. To illustrate that, the results for a 50-50 mixture of logical AND and logical OR are shown in Supplementary Figure 3C: The growth function was fitted best by a power law function. The code for these calculations can be found here:
https://github.com/dankonikolic/Learning-XOR-AND-OR

View publication stats