Noise Robustness and Conscious Supremacy

Ken Mogi
Sony Computer Science Laboratories, Tokyo, Japan
[4]The University of Tokyo, Tokyo, Japan
kenmogi@qualia-manifesto.com

Abstract
It is of interest whether artificial intelligence systems can emulate conscious information processing of the human brain. Here I discuss the aspects of conscious computation leading to artificial consciousness. One of the salient differences between biological neural networks and digital computers is the ubiquitous presence of noise in the former. I outline how noise robustness could correlate with conscious supremacy, where the brain exhibits computational capabilities excelling over classical computers, particularly in the presence of noise. Studying noise robustness and conscious supremacy in biological neural networks would help develop artificial intelligence systems of enhanced computational capabilities, with possible emergence of artificial consciousness, and implications for AI alignment.

Introduction

Artificial intelligence systems have progressed rapidly, and cognitive capabilities of AI have been enhanced remarkably. In that process, comparisons have been made between AI and the human brain, or biological neural networks in general.
One striking difference between biological neural networks and conventional AI is the presence of noise in the former (Shadlen & Newsome 1998, Softky & Koch 1993,Tolhurst et al. 1983). Biological neural networks execute computation in the presence of a substantial level of noise. If artificial intelligence systems were required to operate at a comparable noise level, they would have to implement computational schemes different from those typically employed. Noise robustness (Natarajan et al. 2013, Akhtar & Mian 2018) is the idea that computations can be conducted robustly in the presence of noise, uncertainty, and error. Information theory is embedded in probability theory (Shannon 1948), and provides important framework for robust coding of information in the presence of noise. Reduction of redundancy and the statistical efficiency are important in such a process (Barlow 1961).

Qualia (Chalmers 1995) are a key hallmark of conscious information processing, where seemingly Platonic entities are coded by neural activities full of noise. In general, traits of human cognition are one of the most important constraints for further developments of artificial intelligence (Müller & Bostrom 2016). Biological memory processes are tightly coupled with noise robustness. Declarative memory formation seems to be dependent on perceptual and cognitive processes correlating with consciousness. The dissociation between declarative and semantic memories is crucial in sustainable dynamics of cognitive updating, making it possible for the system to stay away from catastrophic forgetting (French 1999), where exposure to new information could disrupt the accumulated weights of learning.
Artificial consciousness, when and if it materializes, would reflect traits of robust conscious information processing in the presence of noise, and would involve different computational strategies than those employed in conventional approaches in artificial intelligence. Thus, it is interesting to consider the conditions and constraints involved in conscious supremacy realizing noise robustness.
Finally, theoretical and empirical findings on natural consciousness and artificial intelligence have important implications for AI alignment (Russell 2019), an increasingly important field concerning AI ethics, AI safety, and human wellbeing.

Noise in Biological Neural Networks

Compared to artificial intelligence systems based on digital computers, the presence of noise has been acknowledged as one of the important denominators of biological neural networks. Specifically, there are considerable variability in the neural responses in terms of action potentials of individual neurons recorded by electrophysiological measurements (Shadlen & Newsome 1998, Softky & Koch 1993,Tolhurst et al. 1983), which need to be taken into account when considering the salient features of computation in the biological neural networks (Faisal, et al. 2008, Stein et al. 2005).

In addition, neural computations might involve parameters at the cellular level. Cellular computations exhibit comparable degrees of complexity as the network level. Bray (1995) argued that within the cells, proteins and other biomolecules were involved in complex computation in par with neural networks. Transcriptional regulators are involved in cellular computations in the synthetic biology context (Elowitz & Leibler 2000), while biochemical networks can be designed to give functionality (Gardner et al. 2000). There are interesting applications of AI to analyze and possibly design cellular computations (Nilsson et al. 2022), and the concept of artificial biological intelligence (ABI) poses interesting possibilities for further development of computation in natural, artificial, and hybrid systems (Woolfson 2025)

Molecular computation in the cell are under different constraints from artificial intelligence systems realized on digital computers, even more stringent than biological neural networks in some aspects. The low Reynolds number environment in the cell (Purcell, E. M. 1977) is a strict constraint on any molecular computing in the cell, where information coded in states of molecules other than in structure and binding is rapidly lost, ruled by diffusion limited reactions.

Such an environment affects the kind of singularity argument put forward by, for example, Kurzweil (2024), where increasingly sophisticated hordes of molecular machines are hypothesized to facilitate singularity at the molecular level. Realizing molecular singularity as Kurzweil suggested would be difficult without understanding ways to realize noise robustness, as in the case of cell motility (Hunt & Howard 1993, Burgess et al. 2003) and muscle contraction (Lehman & Szent-Györgyi 1975, Herzog et al. 2015) which are conducted efficiently in the presence of noise and low Reynolds number, possibly involving mechanisms of thermal ratchet (Feynman et al. 1966) and principles of information thermodynamics (Sagawa & Ueda 2009).

Consciousness and Computation

There are several models of conscious computation which tries to characterize computational processes unique to consciousness. The global workspace theory (Baars 2005) posits that information processing enter consciousness when its contents are broadcast to various regions in the brain. The integrated information theory (Tononi 2004) starts from the assumption that intrinsic causal relation subserves consciousness.

Models of consciousness do not necessarily address computation as the key ingredients of investigation. Indeed, the hard problem of consciousness (Chalmers 1995) isolates the phenomenological aspects of consciousness as themes of primary concern when it comes to the essence of consciousness, although Chalmers considers qualia to ultimately correlate with functional aspects of computation.

When we consider ways of aligning theories of consciousness with the development of artificial intelligence, computational considerations become of central importance. Some authors (e.g. Penrose 1989) hypothesize that consciousness is capable of non-computable processes, so that conscious computation can execute computation classic computers are not capable of, where understanding is proposed to play an essential role in going beyond the computational capabilities of algorithms, possibly involving quantum gravity. Such an approach, although theoretically interesting, does not necessarily account for the unique capabilities of biological neural networks, due to some difficulties such as the absence of quantum coherence at room temperature. In addition, the distinction between computable and non-computable processes might be academic at best, and not necessarily pertinent for the characterization of the robustness of biological neural networks, especially in the presence of noise.

Thus, we are led to the necessity to consider the characteristics of conscious computation within the domain of computability. Indeed, the watershed between conscious and non-conscious computation would exist within the domain of computability, especially in the presence of noise.

Conscious Supremacy

Conscious supremacy (Mogi 2024) is the idea that biological neural networks might be capable of a set of computations that cannot be executed by classic systems within practical time, although the computation is within the domain of computability. Conscious supremacy is similar to quantum supremacy (Arute et al. 2019) in that the distinction is not on the non-computable nature of information processing, but rather on the practical time and computational resources required to execute computation.

Several aspects of computation executed by biological neural networks are suggested as candidate instances of conscious supremacy: E.g., flexible attention modulation, robust handling of new contexts, choice and decision making, cognition reflecting a wide spectrum of sensory information in an integrated manner, and embodied cognition, would possibly constitute the range of computations exhibiting conscious supremacy.

Mathematical models of computations have been largely developed in the absence of substantial noise. Algorithmic complexity has been discussed without explicit reference to noise, in the mathematical domain of functions (Kolmogorov 1965, Chaitin 1969, Hartmanis & Stearns 1965).

To characterize conscious supremacy, it is important to embrace noise robustness as its central thesis. Specifically, conscious supremacy posits that biological neural networks are able to conduct computation efficiently and robustly, despite the fact that there are substantial noise.

Conscious experience is phenomenologically coherent and integrated within the self-consciousness of the agent. There arises therefore an intriguing question how this is done computationally, given the fact that the biological neural network in the brain is full of noise. Indeed, the nature of action potentials is such that activities of a particular neuron would be virtually invisible from another neuron in the brain separated by a few synapses. An analogy could be made between this situation in the biological neural network and the Greisen–Zatsepin–Kuzmin (GZK) limit (Greisen 1966, Zatsepin & Kuz'min, 1966) in cosmology, which describes how ultra-high-energy cosmic ray protons lose energy rapidly by interacting with photons of the cosmic microwave background (CMB), so that it becomes difficult to observe distant parts of the universe through that range of wavelength. Likewise, regions of the brain is informationally invisible to each other, because of the presence of noise in neural firings. The biological neural network must somehow establish a robust basis for computation against this informational invisibility. That is where conscious supremacy could make a contribution.

There are possibly beneficial aspects of noise for computation. In the process of training, the presence of noise could help the network become more general (Bishop 1995). In stochastic resonance, the presence of noise could enhance signal detection (Benzi et al. 1981), and support computation in biological neural networks in various contexts (Douglass et al.1993, Wiesenfeld & Mos 1995) .

Conscious supremacy in the context of noise robustness explains the origin of consciousness as the adaptation strategy of biological systems in developing robust computation in the presence of noise. Thus, it is different from computational strategies adapted by artificial intelligence systems based on rigid digital memory architecture where invisibility similar to Greisen–Zatsepin–Kuzmin (GZK) limit does not exist.

Noise Robustness and Consciousness

In the evolution of the central nervous system, the presence of noise has been one of the important constraints, and possibly induced consciousness to evolve. In the brain, biomolecules react with each other in an environment where Brownian motions and diffusion limited reactions dominate. It is important to understand the thermodynamic constraints on information processing in such systems.

The Landauer limit states that it takes a minimum energy of $k_B T \ln 2$ to erase one bit of information (Landauer 1961), where k_B is the Boltzmann constant. Advancements in digital computing technology are approaching the Landauer limit at the nanoscale. The biological neural networks operate far above the Landauer limit. However, because of the particular environment within cells adaptation to noise becomes essential. The thermodynamic constraints (Sagawa & Ueda 2009) on computation have implications for the tangible and physical constraints on computation, such as Maxwell's demon, and could have implications for cellular processes, especially in biological motility.

In quantum computation, noise is an essential denominator of what computations are possible. Entanglement with the environment results in decoherence of the qubits. Unlike bits in classical digital computers, qubits cannot be copied (no-cloning theorem, Dieks 1982, Wootters & Zurek 1982). Quantum error correction (QEC, Cai & Ma 2021) has been suggested as a mechanism to rectify the disruptions in computation due to noise in quantum computing, by effectively preserving quantum information without observing the states, which would lead to disruption of coherence. The locally coded information representation can be made robust by coding it non-locally in terms of entangled qubits (Shor 1995). In addition, converting continuous errors into discretized Pauli errors (Knill & Laflamme 1997, Gottesman 1997), would help rectifying the disturbing effects from interaction with the environment.

Likewise, noise is likely to be a crucial denominator in conscious supremacy in the presence of noise. In the evolution of biological systems, consciousness would have provided a framework in which computations could be executed in robust and flexible manners.

Noise robustness needs to be considered over several temporal scales (Fig.1). At the duration of the specious moment in the flow of consciousness, qualia support the phenomenological representation of sensory features in consciousness. The way qualia support conscious computation, although the exact details are still to be elucidated, is likely to share some properties with hybrid systems of symbols and networks (Smolensky 1988, Mao et al. 2019). Indeed, qualia would provide the natural basis for a hybrid system for computation, where the Platonic nature of representation by qualia would provide a symbolic system of representing information about the outside world. In that process, qualia would represent information about the outside world in tangible manners, in an integrated and parallel framework for the conscious agent.

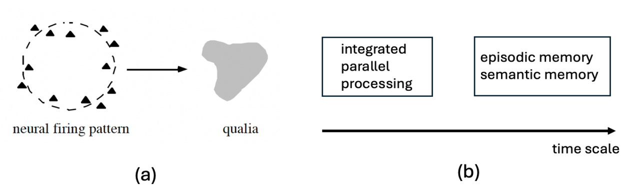

Figure 1: Noise robustness, qualia, and memory.

In longer time scales, episodic memory (Tulving 1972) as distinct from semantic memory (Binder et al. 2009) is important in keeping noise robustness in the buildup of world models, e.g. in avoiding the adversarial effects of catastrophic forgetting (French 1999, Goodfellow et al 2013, Parisi et al. 2019). By storing information about the world in instance-specific terms, explicitly depending on the context of place and time, episodic memory formations can contribute to the long-term refinements of semantic memories, without disrupting the results of former memories.

Episodic memories, in turn, are essentially dependent on the presence of conscious computation for its formation (Tulving 1985), where the agent consciously perceives information about the environment at the specious moment in coherent manners, providing the basis for episodic memory. Remembering past events stored in episodic memories later are mental time travel (Suddendorf & Corballis 1997), in which the agent's conscious experience is temporally linked to experiences in the past. Such linkages between consciousness and memory makes biological neural network computation more flexible and robust.

From such analysis, some conditions for conscious supremacy in the presence of noise emerge.

One, in coding for information in the presence of noise, elements of perception such as qualia would be represented nonlocally, linking neural activities across synaptic connections. In the formation of neurons selectively responsible to certain features (Hubel, & Wiesel 1968, Ramirez et al. 2024), patterns of synaptic connections are important. Response selectivity has been suggested to involve synaptic precision to realize selectivity in spiking responses (Volgushev & Sompolinsky 2002), and any information coded would have to be nonlocally represented.

At present, details of the corresponding principle between response selectivity and qualia, the building block of conscious information processing, is not known. Since qualia would arise from mutual relationships between neural firings, ideas similar to Mach's principle (Barbour & Pfister 1995) might at work (Mogi 1999), where the identity of individuals (qualia) would be formed through relationships between them.

Second, discretization of information in terms of qualia would work as a stabilizer in the conscious computation supported by biological neural networks. Action potentials in neurons are themselves results of discretization invoked at the threshold of membrane potentials in neurons. In addition, discretization on the aggregates of action potentials connected via synapses would work as a stabilizer of computation in the biological neural networks, evidenced by the central position of qualia in the phenomenology of computation conducted by biological neural networks, although the mechanism is still unknown.

Artificial Consciousness

Artificial consciousness is one of the most important research topics to be associated with artificial intelligence research, in terms of AI safety, AI ethics, and general discussions on the practical and philosophical implications of artificial intelligence. For example, metacognition, an integral element of consciousness, might be necessary to make artificial intelligence stable and robust in applications from generative AI (Vaswani et al. 2017, Ho, et al. 2020) to self-driving cars (Yang et al. 2025). The possibility that artificial intelligence systems might be conscious has obvious implications for the ethical treatment of and alignment with AI. Thus, artificial intelligence is of considerable interest not only in terms of philosophical interest, but also in the context of implementing AI in practical contexts.

If, as I have hypothesized, noise robustness was one of the constraints that led to the evolution of consciousness, the current directions of research and development of artificial intelligence systems might not necessarily be aligned with the making of artificial consciousness, in terms of the incorporation of noise as a determining constraint for computation. As a matter of first principle, it is possible that consciousness might arise from computational processes that are based on noiseless processes. However, in terms of understanding consciousness as a biological phenomenon, taking into account noise as an essential constraint for robust computation would seem necessary.

Conscious supremacy in the presence of noise would suggest constraints for artificial intelligence to develop consciousness. Specifically, systems with conscious supremacy would be able to execute computations effectively even in the presence of noise, in ways different from the conventional models of artificial intelligence which typically execute computation in the absence of noise. In the hypothetical construction process, artificial intelligence systems employing conscious supremacy will exhibit a series of unique capabilities in terms of noise robustness, flexibility in computation, and creativity.

It is possible that as the line width of semiconductors used in artificial intelligence becomes narrower, the thermodynamic constraints such as the Landauer limit cannot be ignored any more. However, if we think of artificial intelligence in the presence of noise as a continuation of the conventional approach, there would be more disruptions in place of significant and useful emergence. Alternatively, we may approach the emergence of conscious supremacy in the presence of noise on the layer of simulated system on top of GPUs unaffected by thermodynamic constraints.

It is interesting how higher order properties of consciousness such as free will and self-consciousness could emerge in artificial intelligence systems.

In the compatibilist models of free will, which is the *de facto* mainstream position at present (Frankfurt 1971, Dennett

2003), free will is modelled to arise from autonomous and balanced cognitive processes leading up to one's volition without external coercion. In the compatibilist approach, free will is not an illusion. It is also possible to take a hybrid position where one regards free will as robustly supported cognitively, but also correlates in part with some illusory ideas in the world model (Mogi 2013). Hallucination can arise when the system interprets noise as signal (Fletcher & Frith 2009), and could be involved in the brain's mechanism of interpreting internally generated variability as sources for one's volition, leading to free will.

The cognitive construction of the self is a continuous process of inference about the sensory self (what am I experiencing) and active self (what am I doing), and would depend on the sensory input and internally generated information, which contains noise. In the interpretation of noise, a significant prior would play a pivotal role, and the self can be taken as a framework for the prior (Friston 2005). The self in this sense would play a stabilizing role in conscious supremacy in the presence of noise, and would play an substantial role in the eventual construction of artificial consciousness.

Implications for AI Alignment

Understanding the computational process of natural consciousness and theoretical foundations for artificial consciousness has implications for AI alignment (Russell 2019), an increasingly important aspect of theorizing and developing artificial intelligence, affecting human user experience and AI safety.

From the computational point of view, the division of labor between artificial intelligence and humans would be optimized by focusing on conscious supremacy (Mogi 2024). Specifically, it would be optimum for AI systems to execute those tasks that can be done without involvement of consciousness, while humans would be better off focusing on tasks that essentially require conscious involvement.

As AI systems become increasingly more capable of executing computation from protein folding (Jumper et al. 2021) to mathematical proofs (Hubert et al. 2025), there would be more opportunity for *cognitive supremacy*, where AI systems outperform humans in tasks of high cognitive load, traditionally associated with high IQ, or high g-factor (Spearman 1904). At present, there is no clear reason why cognitive supremacy in this sense could not be achieved without consciousness. Thus, AI alignment would be optimized by AI and humans pursuing cognitive and conscious supremacy, respectively.

A division of labor between humans and artificial intelligence in terms of computational capabilities might also correlate with principles of alignment promoting human wellbeing.

Ikigai (Mogi 2017, Alimujiang et al. 2019), central to human wellbeing, is not simply the perceived goal or target of a particular task, but the experience of life itself. Phenomenological dimensions of life's experiences is important in the constitution and maintenance of ikigai. Recently, the ikigai risk, the possibility that the use of AI might result in reduced ikigai for humans, has been focused (Ziesche & Yampolskiy 2020). Ikigai is typically independent of explicit reward functions, and therefore aligns with the Goodhart's law, which states that "when a measure becomes a target, it ceases to be a good measure." (Strathern 1997).

By delegating cognitive and manipulative tasks to artificial intelligence systems, humans would be able to have more space and flexibility for pursuing ikigai. On the other hand, if the interactions with artificial intelligence constrains the cognitive and behavioral space of humans in such a way that the perceived and experienced ikigai is diminished, that would be detrimental for human wellbeing. In order to enhance ikigai through alignment with AI, it would be beneficial to include an ikigai term in the RLHF (Reinforcement learning from human feedback,) process.

$$\boxed{\max_{\theta}\ \mathbb{E}_{x\sim\mathcal{D},\ y\sim\pi_\theta(\cdot|x)}\Big[r_\phi(x,y)\ -\ \beta\,\mathrm{KL}(\pi_\theta(\cdot|x)\ \|\ \pi_{\mathrm{ref}}(\cdot|x))\Big]}$$

In the standard equation, θ represents the neural network weights, x represents the prompt and context, $\mathcal{D}$is the distribution of prompt, y represents the output of the model, π_θ is the policy to be trained, r_ϕ *(x,y)* is the reward model, π_{ref} is the reference policy prior to RLHF, β is the parameter describing the tradeoff between stability and alignment, and KL is the Kullback–Leibler divergence (Kullback & Leibler 1951) representing the penalty for deviating from the reference policy. The ikigai term would be reflected in r_ϕ *(x,y),* which represents human preference for more or undisturbed ikigai.

Ikigai is closely related to secure base (Bowlby 1988) and attachment (Ainsworth et al. 1978) in developmental psychology, with ikigai serving as a secure base for exploration, where it is maintained independent of external reward functions. In the best scenario, an artificial intelligence system could function as an attachment figure, where human subjects would try to be closer to AI when there is perceived uncertainty and risk. In the strange situation scenario studied in Ainsworth et al. (1978), an infant or child typically tries to make eye contact with, and approaches, an attachment figure (an caretaker, typically a mother or a father). In the artificial intelligence as an attachment figure model, artificial intelligence could take care of humans (Good 1965), as superintelligent guardians (Bostrom 2014).

Noise is an essential factor in AI alignment in the secure base and attachment context, as noise is unavoidable due to the uncertainty in the interaction with the environment. A typical child needs secure base and attachment figure because there are unknowns in the environment. Likewise, humans in the future would need superintelligent AI (AGI, artificial general intelligence, and ASI, artificial superintelligence) as attachment figures providing secure base. In this sense, formation of secure base and attachment figure in interpersonal interactions and humans on one hand, and the AI alignment on the other, could both be regarded as a continuation of the computational principles behind conscious supremacy, where the presence of noise is overcome by nonlocal cognitive processes.

Discussion

Here I argued that consciousness possibly evolved as a means of overcoming noise inherently present in biological neural networks and in the interaction with the environment. I defined conscious supremacy as the ability of conscious computational processes in the biological neural network to realize robust and flexible computation, and suggested two possible mechanisms, nonlocal linking and discretization, as the basis for conscious supremacy realizing noise robustness. Qualia are generated as platonic elements of perception in this process, and would serve as a foundation for noise robustness like in hybrid systems.

At present, the exact nature of computational mechanisms in the proposed scheme is not clear. One criticism would be that computational models with conscious supremacy at its present stage adds nothing to the conventional mathematical formalism of spiking neurons. Another criticism would that computational considerations alone do not solve the hard problem of consciousness such as qualia (Chalmers 1995).

I do not claim to be able to solve the problem of consciousness and methodological difficulties towards building artificial consciousness straightaway with the approach proposed here. I do suggest, however, that noise has been largely overlooked in the discussions of the neural correlates of consciousness (Crick & Koch 1990, Koch et al. 2016), especially in the computational context. Artificial intelligence systems, on the other hand, have been developed on the strength of digital computers where the memories, once stablished, could be preserved stably for a long time, remaining invariant at least for the duration of development and testing of artificial intelligence systems. The synaptic and cellular mechanisms for memory systems in the brain, whether episodic, semantic or otherwise, are very different, where every trace of activity is subject to thermodynamic fluctuations and disturbances.

Embodiment is one of the key issues in understanding consciousness and artificial intelligence, in the coming age of physical AI in particular. The fact that consciousness in biological neural networks are embedded in such a way that there is a substantial level of noise should not be overlooked. It is reasonable to hypothesize that consciousness indeed evolved, at least in part, to address the challenge of noise robustness.

Just picture how, in a human brain, millions of neurons are active at a particular moment, with a lot of variabilities in their activities. Picture, on the other hand, the apparently Platonic perfection of the redness of red, the feeling of the sound of the violin, and the sweetness of a piece of chocolate on your tongue. If one directs one's attention to the discrepancy between the noise in biological neural networks on one hand and the qualia in conscious experience on the other, one realizes that one must somehow bridge these striking different two worlds. Conscious supremacy outlined here could be that bridge, although much work remains to be done.

Conclusion

The key insight leading to the ideas presented here is that in the evolution of the central nervous system, the presence of noise might have been one of the important constraints that induced consciousness to evolve.

The hypothesis of conscious supremacy, where conscious computation is conjectured to bring about robustness in the presence of noise, would hopefully facilitate research into the nature of consciousness and further developments of artificial intelligence, possibly leading to artificial consciousness.

Conscious supremacy would also be helpful in sorting out issues and challenges in AI alignment, contributing to AI safety and wellbeing of human users.

Acknowledgments

I thank my colleagues Takashi Ikegami and Shun Yoshizawa for helpful discussions.

References

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978).

Patterns of attachment: A psychological study of the strange situation. Erlbaum.

Akhtar, N., & Mian, A. (2018). Threat of adversarial attacks on deep learning in computer vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41, 3049–3076.

Alimujiang, A., Wiens, K., Kawachi, I., Aida, J., Tsuji, T., Kondo, K., & Kondo, N. (2019). Association between sense of purpose in life and mortality among adults in Japan: A population-based cohort study. JAMA Network Open, 2(5), e194708.

Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J. C., Barends, R., et al. (2019). Quantum supremacy using a programmable superconducting processor. *Nature €i0574, 505–510. doi: 10.1038/s41586-019-1666-5*

Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Prog. Brain Res. 150, 45–53. doi: 10.1016/S0079-6123(05)50004-9

Barbour, J., & Pfister, H. (Eds.). (1995). Mach's principle: From Newton's bucket to quantum gravity. Boston, MA: Birkhäuser.

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In *Sensory Communication* (W. A. Rosenblith, Ed.), MIT Press, pp. 217–234.

Benzi, R., Sutera, A., & Vulpiani, A. (1981). *The mechanism of stochastic resonance.* Journal of Physics A: Mathematical and General, 14, L453–L457.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex 19, 2767–2796.

Bishop, C. M. (1995). Training with noise is equivalent to Tikhonov regularization. *Neural computation*, *7*(1), 108-116.

Bray, D. (1995). Protein molecules as computational elements in living cells. *Nature 376, 307–312. doi: 10.1038/376307a0*

Bowlby, J. (1988). A secure base: Parent-child attachment and healthy human development. New York: Basic Books.

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.

Burgess, S. A., Walker, M. L., Sakakibara, H., Knight, P. J., & Oiwa, K. (2003). Dynein structure and power stroke. Nature 421, 715-718.

Cai, W., and Ma, Y. (2021). Bosonic quantum error correction codes in superconducting quantum circuits. Fundamental Res. 1, 50–67. doi: 10.1016/j.fmre.2020.12.006

Chaitin, G. J. (1969). On the length of programs for computing finite binary sequences: Statistical considerations. Journal of the ACM, 16(1), 145–159.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences, 2*, 263–275.

Dennett, D. C. (2003). Freedom evolves**.** Viking Press.

Dieks, D. (1982). Communication by EPR devices. Physics Letters A, 92(6), 271–272.

Douglass, J. K., Wilkens, L., Pantazelou, E., & Moss, F. (1993). *Noise enhancement of information transfer in crayfish mechanoreceptors.* Nature, 365**,** 337–340.

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature, 403(6767), 335–338.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience, 9*(4), 292–303. https://doi.org/10.1038/nrn2258

Feynman, R. P., Leighton, R. B., & Sands, M. (1966). The ratchet and pawl. In *The Feynman lectures on physics* (Vol. 1, Chap. 46). Addison-Wesley.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. The Journal of Philosophy, 68(1), 5–20.

French, R.M. (1999). Catastrophic forgetting in connectionist networks. Trends in Cognitive Sciences, 3(4), 128–135.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. Nature Reviews Neuroscience, 10(1), 48–58.

Friston, K. (2005). A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1456), 815–836.

Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. Nature, 403(6767), 339–342.

Good, I. J. (1965). Speculations concerning the first ultraintelligent machine. Advances in Computers, 6, 31–88.

Goodfellow, I. J., Mirza, M., Xiao, D., Courville, A., & Bengio, Y. (2013). An empirical investigation of catastrophic forgetting in gradient-based neural networks. arXiv:1312.6211 [cs.LG]

Gottesman, D. (1997). Stabilizer codes and quantum error correction (Doctoral dissertation, California Institute of Technology).

Greisen, K. (1966). End to the cosmic-ray spectrum?. *Physical Review Letters*. **16** (17): 748–750.

Hartmanis, J., & Stearns, R. E. (1965). On the computational complexity of algorithms. Transactions of the American Mathematical Society, 117(1), 285–306

Herzog, W., Powers, K., Johnston, K., & Duvall, M. (2015). A new paradigm for muscle contraction. *Frontiers in physiology*, *6*, 174.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840–6851.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. The Journal of Physiology, 195(1), 215–243.

Hunt, A. J., & Howard, J. (1993). Kinesin swivels to permit microtubule movement in any direction. *Proceedings of the National Academy of Sciences*, *90*(24), 11653-11657.

Hubert, T., Schrittwieser, J., Guzmán, F., Lample, G., Li, Y., Lee, D., … & Silver, D. (2025). *Olympiad-level formal mathematical reasoning with reinforcement learning*. Nature, 631, 457–465.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583–589

Knill, E., & Laflamme, R. (1997). Theory of quantum error-correcting codes. Physical Review A, 55(2), 900–911.

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. Nature Reviews Neuroscience, 17(5), 307–321.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. Problems of Information Transmission, 1*(1)*, 1–7.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. The Annals of Mathematical Statistics, 22(1), 79–86.

Kurzweil, R. (2024). *The singularity is nearer: when we merge with AI*. Viking.

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, *5*(3), 183–191.

Lehman, W., & Szent-Györgyi, A. G. (1975). Regulation of muscular contraction. Distribution of actin control and myosin control in the animal kingdom. *The Journal of general physiology*, *66*(1), 1-30.

Mao, J., Gan, C., Kohli, P., Tenenbaum, J. B., & Wu, J. (2019). The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4571–4580.

Mogi, K. (1999). Response Selectivity, Neuron Doctrine, and Mach's Principle in Perception. In *Understanding Representation in the Cognitive Sciences: Does Representation Need Reality?* (pp. 127-134). Boston, MA: Springer US.

Mogi, K. (2013). Cognitive factors correlating with the metacognition of the phenomenal properties of experience. *Scientific Reports, 3*, 3354.

Mogi, K. (2017). *The Little Book of Ikigai: The secret Japanese way to live a happy and long life*. Quercus.

Mogi, K. (2024). Artificial intelligence, human cognition, and conscious supremacy. *Frontiers in Psychology*, *15*, 1364714.

Müller, V. C., & Bostrom, N. (2016). *Future progress in artificial intelligence: A survey of expert opinion*. In V. C. Müller (Ed.), *Fundamental issues of artificial intelligence* (Synthese Library, Vol. 376, pp. 553–571). Springer.

Natarajan, N., Dhillon, I. S., Ravikumar, P., & Tewari, A. (2013). Learning with noisy labels. Advances in Neural Information Processing Systems, 26, 1196–1204.

Nilsson, A., Peters, J. M., Meimetis, N., Bryson, B. D., & Lauffenburger, D. A. (2022). *Artificial neural networks enable genome-scale simulations of intracellular signaling.* Nature Communications, 13, Article 3069. https://doi.org/10.1038/s41467-022-30684-y

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. Neural Networks*,* 113*,* 54–71.

Penrose, R. (1989). The emperor's new mind: Concerning computers, minds, and the laws of physics. Oxford University Press.

Purcell, E. M. (1977). Life at low Reynolds number. *American Journal of Physics*, *45*(1). 3–11

Ramirez, J. G., Vanhoyland, M., Murty, N. A. R., Decramer, T., Van Paesschen, W., Bracci, S., Op de Beeck, H., Kanwisher, N., Janssen, P., & Theys, T. (2024). Intracortical recordings reveal the neuronal selectivity for bodies and body parts in the human visual cortex. Proceedings of the National Academy of Sciences of the United States of America, 121(51), e2408871121.

Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking Press.

Sagawa, T., & Ueda, M. (2009). Minimal energy cost for thermodynamic information processing: Measurement and erasure. Physical Review Letters, 102(25), 250602.

Shor, P. W. (1995). Scheme for reducing decoherence in quantum computer memory. Physical Review A, 52(4), R2493–R2496

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *The Journal of Neuroscience,* *18*(10), 3870–3896. https://doi.org/10.1523/JNEUROSCI.18-10-03870.1998

Shannon, C. E. （1948） *A mathematical theory of communication. Bell System Technical Journal* 27, 379-423, 623-656

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *The Journal of Neuroscience, 13*(1), 334–350. https://doi.org/10.1523/JNEUROSCI.13-01-00334.1993

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*(1), 1–74.

Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology, 15*(2), 201–293.

Stein, R. B., Gossen, E. R., & Jones, K. E. (2005). Neuronal variability: Noise or part of the signal? Nature Reviews Neuroscience, 6(5), 389–397. https://doi.org/10.1038/nrn1668

Strathern, M. (1997). *"Improving ratings": Audit in the British University system*. European Review, 5, 305–321.

Suddendorf, T., & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs, 123*, 133–167.

Mogi, K., & Tamori, Y. (1997, September). Creativity and the neural basis of qualia. In *Proceedings of the International Mind II Conference Computational Models of Creative Cognition*. Benjamins Amsterdam.
Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Research, 23(8), 775–785. https://doi.org/10.1016/0042-6989(83)90200-6
Tononi, G. (2004). *An information integration theory of consciousness*. BMC Neuroscience, 5, 42.
Tulving, E. (1972). *Episodic and semantic memory*. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.
Tulving, E. (1985). Memory and consciousness. Canadian Psychology 26, 1–12.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
Volgushev, M., Vidyasagar, T. R., & Sompolinsky, H. (2002). A novel mechanism of response selectivity of neurons in cat visual cortex. Journal of Neuroscience, 22(6), 3038–3050.
Wiesenfeld, K., & Moss, F. (1995). Stochastic resonance and the benefits of noise: From ice ages to crayfish and SQUIDs. Nature, 373, 33–36.
Woolfson, A. (2025). ABI and generative biology: A new paradigm for gene therapy, genome engineering, and engineered cell therapy. Molecular Therapy, *33*(5)
Wootters, W. K., & Zurek, W. H. (1982). *A single quantum cannot be cloned*. Nature, 299(5886), 802–803.
Yang, Y., Liao, H., Zhang, K., & Zhu, H. (2025). Recent advances in interactive driving of autonomous vehicles. International Journal of Automotive Technology, 26(3), 345–362.
Zatsepin, G.T.; Kuz'min, V.A. (1966). Upper limit of the spectrum of cosmic rays. Journal of Experimental and Theoretical Physics Letters. 4: 78–80.
Ziesche, S., & Yampolskiy, R. (2020). Introducing the concept of ikigai to the ethics of AI and of human enhancements. *IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* 138-145.