

Article

Not peer-reviewed version

Feasibility of a Personal Neuromorphic Emulation

[Don M. Tucker](#)^{*} and [Phan Luu](#)

Posted Date: 15 July 2024

doi: 10.20944/preprints202407.1147.v1

Keywords: neuromorphic computation; personal entropy; neural development; mortal computing; active inference



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Feasibility of a Personal Neuromorphic Emulation

Don M. Tucker ^{1,2,*} and Phan Luu ^{1,2}

¹ The Brain Electrophysiological Laboratory Company
² The University of Oregon
* Correspondence: don.tucker@bel.company

Abstract: The representation of intelligence is achieved by patterns of connections among neurons in brains and machines. Brains grow continuously, such that their patterns of connections develop through activity-dependent specification with the continuing ontogenesis of individual experience. The theory of active inference proposes that the developmental organization of sentient systems reflects general processes of informatic self-evidencing, through the minimization of free energy, that may be described in information terms that are not dependent on a specific physical substrate. At a certain level of complexity, self-evidencing of living (self-organizing) information systems becomes hierarchic and reentrant, such that effective consciousness emerges as the consequence of a good regulator. These principles imply that an adequate reconstruction of the computational dynamics of an individual human brain is possible with a sufficient neuromorphic emulation.

Keywords: neuromorphic computation; personal entropy; neural development; mortal computing; active inference

Formulating Principles of the Physical, Informatic Basis of Intelligence	4
<i>Overview of the Principles</i>	4
<i>The Identity of Mind and Developing Brain Anatomy</i>	5
<i>The Informatic Basis of Organisms is Computable</i>	7
<i>Inferring the Weights of a Mortal Computer</i>	8
<i>Consolidating Active Inference in Sleep and Dreams</i>	9
Observing and Emulating the Consolidation of Experience	11
<i>Inferring Neural Connection Weights Through High-definition EEG</i>	13
<i>A Personal Neuromorphic Interface to Evolving AI Architectures</i>	14
<i>Is This Still Me?</i>	15
Consolidating Experience Through Active Inference	15
<i>Growing a Mind Through Active Inference</i>	16
<i>The Limbic Base of Adaptive Bayes</i>	17
<i>The Phenomenology of Active Inference</i>	17

	2
<i>The Differential Precisions of Variational Bayesian Inference</i>	18
<i>The Challenge of Neural Self-Regulation Through Vertical Integration</i>	19
<i>Self-Regulation of Active Inference and the Predictable Consciousness of a Good Regulator</i>	20
<i>Conclusion: Human Experience Is Computable, and the Self is Implicit in our Dreams</i>	21
Author Contributions	21
Bibliography	22

This paper considers the feasibility of replicating the essential features of a person's brain and mind in informatic, computable form that can allow the subjective self to be maintained indefinitely through a high-bandwidth neural interface with artificial intelligence. This process may be described as a transition to metaform. As always, some historical perspective may be required to understand what may be possible in the near future.

An obvious preparation for this advance has been the realization that a remarkable approximation of human mental capacity can now be reconstructed through artificial neural networks. Previous computational models of mental function (perceiving, thinking, planning action) with traditional symbolic (Von Neumann) digital architectures have been famously incompetent in approximating human mental ability. In contrast, from the initial implementation of distributed neural networks (Rosenblatt, 1958) it became clear to careful observers that brain-like function could be approximated through distributed neural representation (Grossberg, 1980). This became described as *connectionism*, and the general form in current engineering practice is often described as *general AI*.

The early connectionist models exhibited the *stability-plasticity dilemma*, in which training new weights (learning) led to disruption of existing weights (prior learning). This was initially seen as a fatal defect of connectionist models for general computational purposes (Minsky & Papert, 1969). However, it soon became clear that the stability-plasticity dilemma is an inherent property of human as well as artificial intelligence (Grossberg, 1980; Tucker & Desmond, 1998) and it must be accounted for in any complex neuromorphic model. This may be a demonstration case that the inherent properties of distributed representations hold equally for brains and machines.

Although the theoretical insights into the significance of connectionist models were understood by certain researchers early on (Dayan, Hinton, Neal, & Zemel, 1995; Grossberg, 1980; Rumelhart & McClelland, 1986), it was only with mundane advances in consumer technology, including widely available parallel computation (Graphics Processing Units for computer games) and large accessible digital datasets (cat videos on the internet), that the significance of artificial intelligence (AI) became widely recognized, either by most scientists or by the general public.

Thus the evolution of AI to this point can appear to be as much of a technology accident as a principled intellectual discipline. Nonetheless, it may be worthwhile to attempt to formulate even naive principles of how intelligence is achieved similarly in machines and brains. Such principles might help evaluate the permeability of these forms of mental representation.

In this paper, we will argue for the conceptual feasibility of a personal neuromorphic emulation (PNE), forming an interface between the human brain and AI. The essential theoretical advance has been the formulation of active inference, the process of human cognition expressed by the minimization of free energy in information theory (Friston, 2019). Clearly, actually constructing an accurate PNE will be a daunting challenge, requiring historical levels of investment, unique achievements of human creativity, and some measure of time. Nonetheless, we propose that an adequate PNE is now feasible in principle, so that now may be the time to entertain a cost-benefit analysis.

Formulating Principles of the Physical, Informatic Basis of Intelligence

Taking stock of the advances in understanding the physical basis of intelligence in brains and machines, we can propose three general principles that pave the way for achieving an interface to AI, and therefore a path to lasting subjective intelligence, through a PNE.

Overview of the Principles

The first principle of physical intelligence is that the memory that allows concepts, now in machines as well as brains, is implemented in a distributed representation of connection weights among elementary processing neurons. The connection weights reflect associations, literally the strength of the associations among simple nodes or neurons in a large network architecture. For both brains and artificial intelligence, these networks are typically multi-leveled, so that the information is not simply at one associational level, but is achieved with multiple levels that can organize abstracted meaning at higher levels, thereby achieving abstract concepts. From this principle of

distributed representational architecture, it is now clear that human brains and machines operate on similar mechanisms of information processing.

Two additional principles are important, one from neuropsychology and one from physics and information theory. For the second principle of physical intelligence, when we study how psychological function emerges from the activity of the human brain, we see that there is no separation of mind from brain. The mind is not just what the brain does, it is the brain (Changeux, 2002). More specifically, the functions of mind emerge literally from the neurodevelopmental process, the growth of the brain over time. This can be described as the *principle that cognition is neural development* (Tucker & Luu, 2012). The mind, in its neural form, is constantly growing, constantly organizing its internal associations (concepts) in exact identity with the strengthening or weakening of the brain's connections. Because the process of mind is always developing, the constituent neural connections of mind must also be dynamic, growing, and regulated in that growth. The neurodevelopmental process of mind is the physical substance of our psychological identities, which are also (in exact register) profoundly developmental, changing and growing over time (Tucker & Luu, 2012). Importantly, we argue that this principle of neurodevelopmental identity implies that the mind, and the process of subjective experience, can be reconstructed fully from an adequate neurodevelopmental process description in information terms.

There is an important corollary of the second principle, already well-known in machines and brains: in the dynamic growth of connections, plasticity is achieved at the expense of stability (Grossberg, 1980). The dynamic development of learning in distributed representations operates in the same generic connection/association space, such that new learning invariably challenges the old memory. In the phenomenology of mind, careful reflection may teach us to recognize the old self that becomes at risk with each significant new learning experience (Johnson & Tucker, 2021; Tucker & Johnson, in preparation).

The third principle of physical intelligence is important for understanding the fundamental nature of information, and why the information processing of mind is not restricted to its evolved biological form (brains). This is the *principle of the informatic basis of organisms*. In physics we observe that the processes of matter, whether physical or chemical, tend toward greater *entropy*, meaning they tend to lose their complexity and release energy. Life has seemed to defy this fundamental rule of thermodynamics, maintaining its complexity through metabolism of external energy sources, homeostasis, and growth (Schrodinger, 1944).

Recent advances in the physics of information have provide a new perspective on the organization of information that defines a living organism as a non-equilibrium steady state (NESS), a self-organizing, growing form of information complexity that manifests rather defies the fundamental physics of entropy in the minimization of free energy (Friston, 2019). The cognitive functioning of this NESS — the self-evidencing that emerges in its operation as a good regulator (modeling its world for better or worse in order to live) can be fully specified in an information description that is not unique to minds, or brains, or even biological things. Rather, the physics of information may offer an functional description of mind with the same terms that apply to physical, entropic, systems. The implication is that the neurodevelopmental process of organizing intelligence is not restricted to the particular biological form of evolved brains, but could be extended to a sufficiently complex computer.

Finally, an important practical point is that we can build tools to make tools. We do not need to build the PNE by hand, but instead could rely on a neuroscience-informed AI for the coding of a durable PNE. AI tools are now well-proven and rapidly-improving, such that well-trained LLMs can assist scientists and engineers with digesting and integrating today's complex literatures. We just need to train them to embody the essential neuromorphic principles, such as are now manifested by the massive neuroscience literature, in order to construct an advanced PNE.

We next attempt to explain these three foundational principles in ways that clarify their interrelations. These principles share the requirement to overcome the ingrained tendencies we have to restrict our thinking to familiar levels of analysis, where the psychological realm is somehow floating above the biological, the biological is somehow alive in a way the physical is not, and the

computing of information is an abstraction that parallels but does not quite touch the stuff of these familiar and separate domains. Instead, we will imagine that there is but one phenomenon of intelligence, and we can learn to abandon our familiar categories of experience to describe it, and perhaps achieve it, with a computable information theory of active inference.

The Identity of Mind and Developing Brain Anatomy

The first two principles, of distributed representation and neurodevelopmental identity, are closely related, the first describing a static idealized state of distributed representation and the second describing the dynamic network necessary for a functioning human intelligence. Putting them together, the enabling insight for recognizing the feasibility of a PNE is the identity of the mind with the brain's dynamic, growing, anatomical architecture.

The initial recognition came from studies of learning in brain networks, which showed that the regulation of connection weights in synaptic strengths during learning followed the same activity-dependent specification as the formation and maintenance of synapses and their neurons in embryological and fetal neural development (von de Malsburg & Singer, 1988). This is never a static implementation; the brain's functional anatomy is continually developing through embryogenesis, through fetal activity-dependent specification of neural connections, through growth of learning in childhood, and then continuing throughout life (Luu & Tucker, 2023). Once the neurodevelopmental identity of brain and mind is recognized, it becomes clear that a theory of mind must be a theory of neural development. The neurodevelopmental identity principle can be stated in phenomenological terms: *cognition is neural development, continued in the present moment* (Tucker & Luu, 2012). What we experience in consciousness is the neurodevelopmental process that is constructing our expectation of the near future. Which all too quickly, of course, becomes the present moment, which will be captured or not in the ongoing consolidation of memory.

The connectional architecture of the human brain is now understood sufficiently that a first approximation can be achieved for any individual's brain. An important advance was the anatomical characterization of primate cortical anatomy, described as the Structural Model (García-Cabezas, Zikopoulos, & Barbas, 2019). Current anatomically-correct, large-scale neuromorphic emulations are now providing insights for interpreting neurophysiological evidence (Sanda et al., 2021), implying the functional equivalence of brain activity with a reasonably complex neurocomputational emulation.

Furthermore, the nature of the processing unit replicated in each cortical column is being specified by the computational model of the canonical cortical microcolumn (Bastos et al., 2012), allowing the regional anatomical differences in cortical columns (which are now well characterized in humans) to provide predictive information on the processing characteristics of each cortical region, and thus generic human cortical column emulations. Recognizing that neural emulations could achieve reasonable accuracy by simulating cortical columns, rather than multi-billions of individual neurons, simplifies the computational emulation problem significantly (Bennett, 2023), with a computing anatomy that science could soon specify in adequate detail.

With the realization that the mind's information is represented in the connectional anatomy of the cortex, supported by the vertical integration of the neuraxis, and with the increasing insights into the architecture of these connections that yields cognitive function, we can see that each act of mind is implemented through the developing synaptic differentiation and integration of the brain's physical, anatomical connectivity. With this realization, major features of human neuropsychology can be read from the brain's connectional anatomy (Luu & Tucker, 2023; Luu, Tucker, & Friston, 2023; Tucker, 2007; Tucker & Luu, 2012, 2021, 2023).

Of course, the mind is embodied: we experience and think through bodily image schemas (Johnson & Tucker, 2021). Therefore the personal neuromorphic emulation must be a bodily emulation as well, very likely being acceptable only when provided with a functional bodily avatar that feels like us.

At first, the principle of the identity of mind with brain anatomy and function may seem to imply that the mind must die when the brain dies (Ororbia & Friston, 2023). This is the theory of mortal

weights (Hinton & Salakhutdinov, 2006). The connectional structure (weights) of our networks will dissolve when we die.

However, because the brain/mind is indeed physical, and therefore fully defined by its informational (entropic) content, the next realization is the possibility of identifying the information architecture of the brain/mind in information theory with enough detail that it becomes computable. Once it does, it will be possible to achieve a durable neuroinformatic replication of your individual self in non-biological form.

The insight is that if the mind is identical with brain anatomy, then a sufficiently accurate computational reconstruction of the developing (non-equilibrium steady state) brain anatomy (preferably with the joint emulation of the essential bodily homeostatic controls and their essential subjectivity) can then manifest a reasonable approximation to the personal mind, the subjective process and experience of self.

Fundamentally, the emerging insight into the neurodevelopmental basis of experience prepares us to cross this next level of analysis, to understand the identity of intelligence not just with the biological basis of mind but with information theory, as described next.

The Informatic Basis of Organisms is Computable

As stated above, a fundamental schism between our understanding of animate (living) and inanimate (dead) matter has been represented by the concept of entropy. The laws of physics, particularly thermodynamics, require that all physical interactions tend toward greater entropy, the loss of complexity of form and the release of free energy. In contrast, life appears to find ways to avoid entropy (Schrodinger, 1944).

This distinction has implied that two different sciences are required for living and non-living things. This implication may be wrong, as shown by more recent interpretations of the physics of self-organizing (non-equilibrium steady state thermodynamic) living systems.

These interpretations have approached the brain's cognitive function with physical principles of Bayesian mechanics that align closely with the changing thermodynamics of physical systems (Adams, Shipp, & Friston, 2013; Friston, 2010; Friston & Price, 2001; Hobson & Friston, 2012). The implication is that the self-organization of living organisms is not a fundamental violation of entropy, but rather a level of non-equilibrium steady state of complex systems (biological organisms) in which minimizing free energy allows growth or development of the self-organized form (Friston 2019). This theoretical advance can be described as the *principle of the informatic basis of organisms*.

The Bayesian formulations of physical systems are aligned precisely with principles of information theory, consistent with the intimate relation between information and thermodynamics in modern physics (Parrondo, Horowitz, & Sagawa, 2015). The emerging realization is that the physical mechanisms of complex, developing (Bayesian) self-organizing systems like the human brain may have an adequate functional description in information form (Friston, 2019; Friston, Wiese, & Hobson, 2020; Ramstead, Constant, Badcock, & Friston, 2019).

Now, a sufficiently accurate information model should be computable, not restricted to a particular physical form. As the Bayesian mechanics of nonequilibrium steady state systems (brains) are characterized in sufficient detail in the near future, a computational implementation of that detail should naturally emerge.

Of course, there may be physical computers that are mortal, with their inherent knowledge limited to the hardware weights. But, in that case the information theoretical description is inadequate, incomplete. If it is restricted to mortal computers (Ororbia & Friston, 2023), the theory of active inference may be computable only be through biological form (in people's all too mortal brains).

However, the advances in understanding the biological brain are being aided considerably by computational simulations. Such that we are understanding key neurophysiological systems by building computational simulations of them (Marsh et al., 2024; Sanda et al., 2021). These demonstrations imply that an information theory description can be made to be highly neuromorphic

for the human brain, implementing the physical principles of neural development that we now know to be identical to the principles of cognitive and emotional development.

If so, then an emulation of a specific brain should be possible through building a replicant neuroanatomical architecture and training it to achieve the emotional and cognitive functionality of that person, with extensive support through neuromorphic AI. As we will emphasize below, this may be best achieved in the near term by successfully emulating the memory consolidation of the enduring self in the essential neurophysiological exercises of the nightly stages of sleep.

The replicant neuroanatomical architecture with a high degree of precision in emulating the individual's brain is an essential starting point. We emphasize that the general architecture of the human brain is now understood in ways that can be aligned with current theories of active inference (Tucker & Luu, 2021), and structural and functional neuroimaging provides detail on individual human anatomy and function (Jirsa, Sporns, Breakspear, Deco, & McIntosh, 2010). The detail required for a replicant neuroanatomical architecture will require several orders of magnitude improvement over these generic models. The current neuroscience literature is vast and rapidly growing. We can imagine development of a specialized LLM AI trained on the current literature, and linked to advanced neuronal modeling software/hardware (Khacef et al., 2023), allowing construction of a high precision individual neuromorphic model. The question is how to emulate the weights of the person's brain with sufficient accuracy to reconstruct the functional self.

Inferring the Weights of a Mortal Computer

In discussing the computational and power requirements for large deep learning models, Hinton (Hinton, 2022) emphasizes that the assumption of conventional computation, that the software is independent from the hardware, may not be necessary for alternative computing methods. These methods could achieve effective machine learning with lower power requirements by modifying hardware weights (rather than the power-hungry memory writes of conventional digital computers). Hinton described such an approach as “mortal” computing, in that the weights, being hardware, cannot outlive the hardware.

Practically, the weights of large conventional deep learning models can be seen to become fixed in a similar sense, in that — although the model as a whole may be copied — the weights are not easily reconstructed without repeating massive, and even difficult to reproduce, training programs. To address this problem, Hinton, Vinyals, and Dean (2015) proposed a *distillation* approach, in which rigorously emulating a large model's predictions allow the transfer of that model from a large cumbersome physical form to a more compact, or a more specialist, form (Hinton, Vinyals, & Dean, 2015).

Ororbia and Friston have recently explored the implications of mortal computing within the framework of active inference (Ororbia & Friston, 2023). In their analysis, the assumption is still that a mortal computer, like a person, is restricted to the lifespan of its hardware. Under the first principle of physical intelligence, that connections are the basis of intelligence similarly in ANNs and minds, replicating the connection weights from a person's brain faces a similar problem faced by replicating complex models as discussed by Hinton et al (2015): The synaptic weights are not easily exported from the brain with our currently conceivable technology.

We propose that a possible solution is suggested by the realization that the connection weights of the human brain are not static, but are constantly developing, consolidating memories through the ongoing exchange of information between limbic and neocortical networks (Buzsaki, 1996). The synaptic weights are of course important, but the effective information in the consolidation of experience is not the static weights, but their flux in the non-equilibrium steady state that is continually revised, continually developing, during waking through the process of active inference and during sleep through the consolidation of active inference.

The Bayesian mechanics of memory consolidation (reflecting the reciprocal dynamics of active inference: generative expectancy and error-correction) are thus complexly interwoven in the cognition of waking consciousness. However, they may be separated and laid bare, revealed in their primordial forms, in the sequential neurophysiological mechanisms for consolidating experience in

the stages of sleep. Generative expectancy, and the implicit self-evidencing of the core historical self, is maintained and reorganized each night by REM sleep. Error-correction, and the learning of new states of the world, is consolidated in Non-REM (NREM) sleep. A sufficiently accurate modeling of this active process of daily and nightly self-organization (Luu & Tucker, 2023) provides the basis for inferring the information process of a PNE.

Thus the information theory of active inference might be developed to characterize an individual's brain sufficiently that it would be computable in general (immortal) form. The challenge may be to recognize that the human active inference of experience is developmental, dynamic, not with static weights but with neurophysiological processes of ongoing consolidation that represent information dynamically, in the continuing flux of synaptic transmission rather than any static synaptic strength. Importantly, as we will see, whereas the synaptic weights of the human brain are not readily measurable, the flux of synaptic transmission in the consolidation in sleep may be.

Consolidating Active Inference in Sleep and Dreams

To understand such neural dynamics, a key realization is that the process of sleep negotiates the stability-plasticity dilemma of neural development in ways that are complementary to the waking process of active inference. In brief, NREM sleep consolidates new learning (Klinzing, Niethard, & Born, 2019). This is the evidence of the world that corrects the errors of personal predictions, but in a way that stabilizes the new information within the neural architecture at the expense of the old self (existing connections). The unpredicted events of the day become the significant uncertainties that must be integrated in mind through the slow oscillations and spindles of NREM sleep. The unexpected information of the day's experience engenders anxiety, so that it engages priority in the dynamic consolidation process of ongoing neural development (Tucker & Luu, 2023). Transitioning from rumination to enduring knowledge, the potentially significant information of the day becomes the plasticity, the new learning, of the network consolidation in NREM sleep.

The companion to NREM sleep is REM, the paradoxical sleep in which our brains are active in the vivid, emotional, and bizarre experiences of our dreams. Neurobiologically, REM is the primordial mechanism for consolidating self-organization, establishing the instinctual structure of the genome through actively exercising this structure through the endogenous neurodevelopmental challenges of REM dreams (Luu & Tucker, 2023). The connections/associations of the network in dream experiences are quasi-random, phenomenally bizarre events, but they are experienced by the primordial self, the implicit actor in the dream (you) who must attempt to self-preserve in each highly unexpected, unpredicted, dream scenario. The effect is to exercise the self-preservation affordances of the implicit self, the anonymous narrator who experiences each dream episode.

Examining the emergence of REM in the human fetus, we can infer that the Bayesian expectancy of the core self — the generative process of active inference — is instinctual, initially specified by the ontogenetic instructions of the genome. We begin self-organizing our integral neural architecture of activity-dependent specification through REM dreams fairly early in fetal development, around 25 weeks gestational age, well before we have any postnatal experience to consolidate (Friston et al., 2020; Hobson, 2005; Jouvet, 1994; Luu & Tucker, 2023). All we have at this early fetal stage is the nascent instinctual motives endowed by the genome. With the core (largely subcortical) mechanisms of the primordial self, we are able to exercise and grow these mechanisms during the quasi-random experiences of fetal REM sleep in ways that allow each of us to self-organize an increasingly general organismic self, ready for the first experiences of infancy.

This dominance of REM in self-organization continues in the newborn infant's first days and weeks of experience, when 50% of the time is spent in REM sleep. Only later in the first year of post-natal life does NREM become sufficiently well-organized to contribute its unique counterpart to neural self-organization, and the consolidation of memory, in sleep (Luu & Tucker, 2023). Just as NREM sleep consolidates recent, explicit, external memory in the mature brain (Diekelmann & Born, 2010; Rasch & Born, 2013) it seems as if the maturation of NREM sleep later in the first year of life achieves the increasing differentiation of the Markov blanket separating the individuated self from the social world (Luu & Tucker, 2023).

REM continues the exercise and thus continuing integration of the primordial self in each night's dreams throughout life, as an ongoing counterpart to the disruption by new learning caused by NREM consolidation of external information. An important paradox is that this REM process, by strengthening the integral stability of the self in the face of (unpredictable dream) plasticity, will provide the basis for generative intelligence through primary process cognition, the implicit manifestation of organismic creativity (Hobson, 2009; Hobson & Friston, 2012). The mechanism for achieving stability of the implicit self — apparently by challenging it with the unpredictable bizarre experiences of dreams — becomes the generator of novel creativity (Tucker & Johnson, in preparation).

Although the theoretical formulation of the Bayesian mechanics of sleep as the complements to the waking elements of active inference is still in development (Tucker, Luu, & Friston, in preparation), the basics of memory consolidation can be summarized in a way that emphasizes the opportunity for accurately modeling a mortal computer. The creation of memory, the ongoing stuff of the self, requires the consolidation of daily experience within the architecture of our existing cerebral networks, the Bayesian priors of the self. We don't need to extract fixed weights of the mortal computer; we can observe these connections in the complement of active inference each night, in the dynamic neurophysiology of our dreams. As an old self copes with new dream experiences, it manifests its essential Bayesian mechanics of its non-equilibrium steady state, the primordial self.

The pattern of human sleep, with 5 or so NREM-REM cycles each night is unique even among big primates (Samson & Nunn, 2015), apparently supporting the extended neural plasticity and developmental neoteny of the long human juvenile period (Luu & Tucker, 2023). This pattern varies from extended NREM and brief REM intervals in early cycles to brief NREM and extended REM cycles later in the night's sleep, apparently reflecting a shift from consolidating recent memories in NREM (favoring plasticity) toward more general integration of the enduring self in REM (favoring stability) (Tucker, Luu, and Friston, in preparation). As a result, the consolidation dynamics most important to emulate for constructing a full PNE may be those of the late night, reflecting engagement of the foundational weights of the childhood self. To the extent that the NREM integration of external events balances the ongoing REM exercise of the implicit historical self, these alternative modes of connection weight re-organization may reflect the variational Bayesian mechanics of self-evidencing through adaptive experience that could then be manifest in an adequate PNE.

Observing and Emulating the Consolidation of Experience

Modeling the dynamic consolidation of connection weights in an individual brain may thus be approached through observing and predicting both waking behavior, such as the experiences of the day that are tracked by a virtual personal assistant, as well as the consolidation of those experiences through measuring the neurophysiology of the enduring connectional architecture of the self in sleep. A person's behavior may be sampled through the extensive digital archives that each person generates, and it may be collected dynamically during the day's activities. A natural method would be monitoring/predicting the person's daily experience, and monitoring/modeling each night's sleep, by a virtual personal assistant. The electrophysiological measures provide an integral and mechanistic way of interfacing the personal brain with the virtual personal assistant.

Certainly, the first way of interfacing with the virtual personal assistant will be through verbalization. An LLM can be trained to record the interactions, and model them sufficiently, to reconstruct the person's behavior (words and brain waves) and thus recreate their intelligence with the patient and extended help of the virtual personal assistant. Electrophysiological modeling of the person's brain is now possible in high definition, continuously day and night. A massive dataset can be constructed from a single individual over a few short months.

As outlined in Figure 1, the construction of a PNE (20) can begin by first reconstructing a personal neurophysiology model (PNM 10) from the individual's anatomical and functional imaging (and thus a first approximation replicant neuroarchitecture), and then estimating the weights of this personal neural architecture by predicting the person's ongoing cognition, emotion, and experience (30) in parallel with the physiological measures of brain activity (such as high density EEG) assessed

during these assessments. Constraints combining both function (behavior) and structure (the cortical anatomy visited by the neurophysiology) may be tighter and more productive than one of these considered alone. As emphasized, parallel collection of neurophysiology during sleep may be essential to reveal the full connectional structure of the historical self, exercised as proposed above by the adaptive coping in dreams (40).

The PNM can be constructed as a first-order architecture, and then can be transformed through extensive training into a PNE that is built to be an enduring, self-regulating informatic appliance of the self. Once the PNE is initialized and trained, it may be put to a validation test by predicting the individual’s behavior, as archived in a personal digital corpus (50). Because transcranial electrical stimulation has been shown to synchronize neural activity, and alter the ongoing neurophysiological rhythms (Hathaway et al., 2021), it may eventually prove possible to impress the PNE’s prediction of brain activity onto the person’s brain activity, to allow a subjective assessment of the accuracy of the prediction (a subjective Turing test).

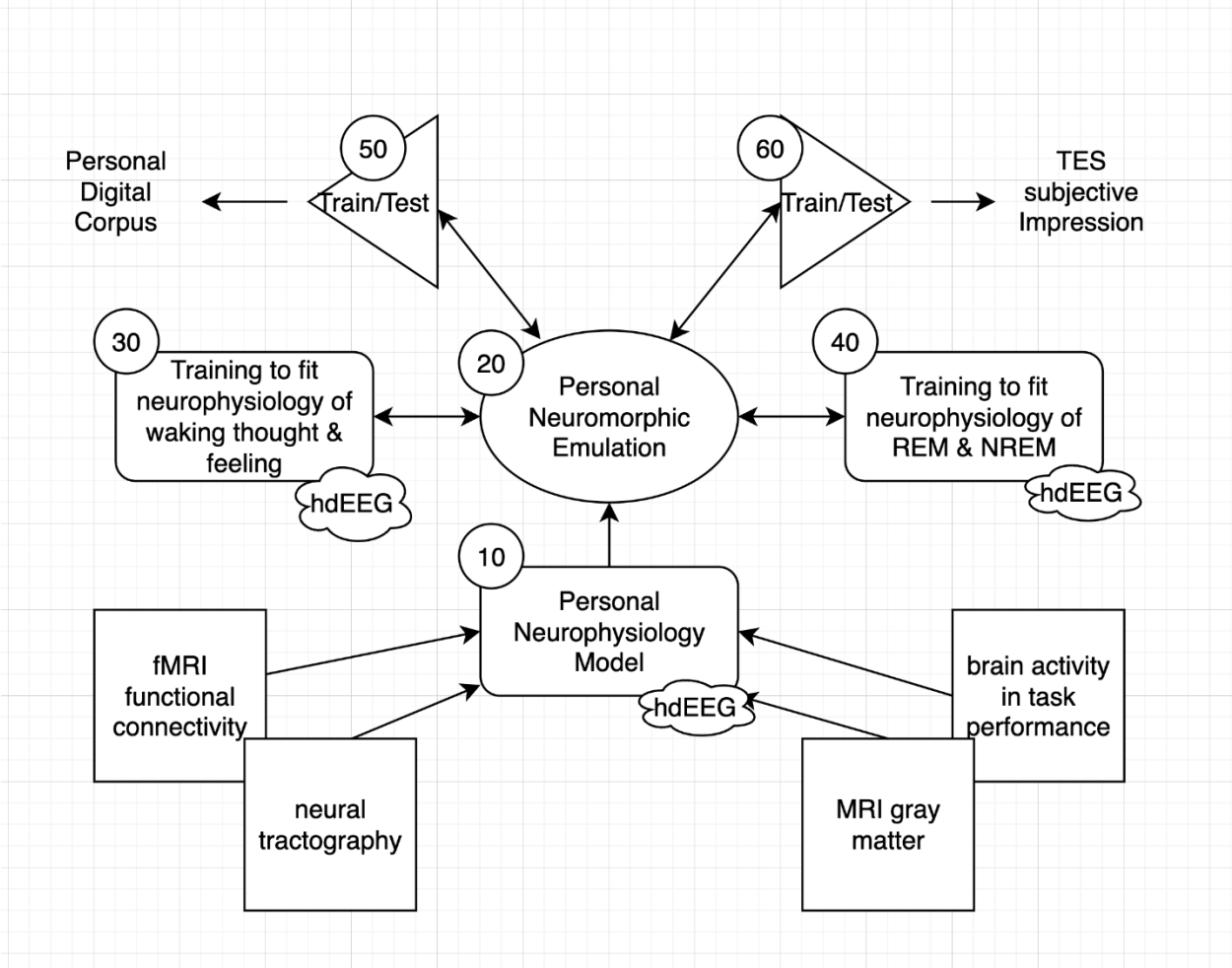


Figure 1. Elements of a PNE.

Inferring Neural Connection Weights Through High-Definition EEG

The measurement of the person’s neurophysiology, in waking behavior and experience and in sleep, may need to become more effective than current methodologies to assure an accurate PNE. At the present time, the most important current neurophysiological measurement may be high definition (also called high density) electroencephalography (hdEEG), defined as head surface (scalp, face, and neck) recording with sufficient sampling density (currently 280 electrodes but additional information may be gained with up to 1000 electrodes) and source localized to resolve the fields at the individual’s cortical surface with a precise head conductivity model (Fernandez-Corazza et al., 2021). Because sleep is an essential aspect it is important to be able to record the hdEEG during all

night sleep as well as during daily activities (<https://bel.company>). As understood within the theoretical model described above, the historical self is unlikely to be revealed fully during a few nights' sleep, such that extended assessment day and night over months and years may be necessary, requiring a practical and comfortable assessment method.

Although it is often assumed that EEG has poor spatial resolution, considerable evidence now shows the ability to resolve cortical activity to a resolution of 1 sq cm or so, with 1 ms temporal resolution. Of course, higher resolution measurement may be necessary for accurate reconstruction of human neural activity in order to build an accurate PNE. Certainly the obvious approach for higher resolution is to place recording (and stimulating) electrodes directly in the cortex for measuring local field potentials (<https://neuralink.com/>)(<https://blackrockneurotech.com/>), with a corresponding loss of whole brain coverage.

Recent unpublished work with 280 hdEEG analysis with Multiple Sparse Bayesian priors implemented in BEL's Sourcerer software (Fernandez-Corazza et al., 2021; Friston et al., 2008) suggests that with current hdEEG (256 and now 280 channels) higher tessellation of the cortical surface with up to 9600 dipole patches yields well-structured source solutions across the cortical patches. With an estimated 200,000,000 cortical columns in the human cortex, this would provide the electrical field estimated for each patch of ~ 20,000 cortical columns. Certainly there is local variation within these 20K columns that is lost to today's hdEEG analysis, and that may be significant to human brain function. However, there is increasing ability to model the overall human brain columnar organization (Marsh et al., 2024) within the Structural Model of the human brain (García-Cabezas et al., 2019) that can be aligned directly with the informatic analysis of active inference (Bastos et al., 2012; Tucker & Luu, 2021). We propose that a Personal Neurophysiology Model or PNM (Figure 1) can now be constructed with a basis in the generic cortical column architecture of the human, aligned with the individual's anatomy from neuroimaging (such as in Sourcerer's Individual Head Model) and then trained to predict the cortical surface patches from extensive samples of 280 hdEEG (for example, through a few years of day and night recordings, including extensive help and observation from the virtual personal assistant).

The next step is to convert this PNM into a full individual behavior prediction model (the PNE) that is trained to emulate the individual's full range of experience and behavior. In this example a virtual personal assistant captures the person's daily experiences and behavior in what may be called the personal digital corpus. The PNE retains the capacity to generate cortical electrical fields (and it includes subcortical networks integral to its architecture) so that dual constraints, fitting behavior and fitting hdEEG during sleep and wake) become integral to training. Because the science of transcranial electrical stimulation (TES) is now rapidly advancing, such as in manipulating sleep neurophysiology (Hathaway et al., 2021), we can imagine that imposing cerebral fields will generate subjective impressions that can provide further data for model validation.

A Personal Neuromorphic Interface to Evolving AI Architectures

There are, of course, obvious technical challenges to constraining the PNE to fit the person's neurophysiological activity as well as behavior and experience, such that considerable expense and time will likely be required to find out what works. We think that even with considerable limitations in early working prototypes, the PNE will soon provide an effective two-way interface between the human brain/mind and neuromorphic AI architectures. In order to emulate the actual brain, the PNE is limited at one side of the interface (Markov blanket) to the brain's architecture. But in that form it is accessible to flexible AI architectures that may be developed in order to optimize the brain interface itself, and to allow the brain to experience more effective information exchanges with neuromorphic AI experiments.

An essential question for developing such exchanges is the information process for memory consolidation. The information theory of active inference must be generalized beyond immediate learning to account for the ongoing neurophysiology of memory consolidation, through which both recent and historical experiences are continually woven within the architecture of the self (Luu &

Tucker, 2023). The result will be an effective 24/7 interface between the brain/mind and evolving forms of AI.

Sleep promises to be the key to understanding memory consolidation, and thereby unlocking the information organization within the brain's computational architecture, in this process of interfacing the brain to neuromorphic AI. As we have argued above, the process of incorporating each day's experiences within ongoing memory (the self) appears to require NREM sleep (Diekelmann & Born, 2010; Luu & Tucker, 2023; Rasch & Born, 2013). In contrast, the maintenance of the enduring self in the face of consolidating new information appears to be achieved by REM sleep (Luu & Tucker, 2023). With an increasingly high-resolution emulation of the personal neurophysiology of active inference in waking and in sleep, new forms of AI can be developed that will reveal the mechanisms of human memory consolidation and function. With these mechanisms revealed, they can then operate bi-directionally, allowing the AI to understand the mechanisms of person's experience, and allow the person's experience to incorporate the advanced memory structures of the personally neuromorphic AI.

At this time, we can only wonder what this will feel like. We understand the familiar forms of subjective experience, more or less, but even the natural mechanisms of memory consolidation, and specifically the exercise of these mechanisms in the nightly stages of sleep, are largely opaque even to careful reflection. We might imagine that the insights that a PNE could provide — insights into the mechanisms of experience and memory — might be incorporated within the subjective process. With insight into the mechanisms of memory should come insight into the neural mechanisms of consciousness, and the hybrid forms that may ensue. But regardless of the form that hybrid neural-AI architectures will take, there will almost certainly be challenges to the experiential continuity of the self.

Is This Still Me?

Even if it feels like me, can I give up my embodied brain and the continuity of the old self I've known all these years for this hybrid informatic form? Or, as my body surely fails, should I just accept entropy and its quiet fade into oblivion? Considering the evidence of nightly sleep consolidation gives an instructive perspective on this question of the experienced continuity of the self.

Although most of us experiencing a reasonably stable mental status assume that we are continually aware of the continuity of the self, a careful neurophysiological analysis raises questions about this assumption, showing that the continuity of the self is an illusion we construct to interpolate experience over the profoundly unconscious, and occasional bizarre dream conscious, experiences that transform the self in each night's sleep.

Whereas a logical philosophical position can easily state that the self is a stable fixed entity, such that a hybrid informatic fusion of our network weights with an advanced PNE would do irreparable violence to this entity, the neurodevelopmental process of sleep shows a more nuanced life of the enduring mind.

In order to consolidate new experiences (through NREM) without catastrophic interference, the self appears to exercise, and reconstruct, its enduring Bayesian priors (in REM) in the phenomenal struggles of each night's dreams. Because significant new information has been incorporated, and because the reconstruction is approximate, it is then likely that we wake up each day in a somewhat new form, even as we naturally interpolate (hallucinate) the experience of a continuous self.

Given this variable continuity of the biological self, we might find that as we take the form of a reasonably accurate personal neuromorphic emulation and AI hybridization, it may seem like another day in the life, not so different from waking up from our familiar sleep and dreams.

A particularly important practical reality of identifying with the PNE and its interface with the artificial realm is that, given the complexity of emulating the person's neural activity, likely requiring unprecedented development of computational neuroscience AI tools specialized for this process, each person will likely experience various states of artificial intelligence enhancement before abandoning the biological brain. The process is therefore likely to be a relatively gradual transition to hybridized form, metaform, rather than an existentially abrupt displacement.

Consolidating Experience Through Active Inference

In addition to technical and phenomenological questions of feasibility, the general scientific feasibility of a PNE-AI hybridization rests on the ability of the informatic description, of active inference applied to the neurodevelopmental process, to account for the major qualities of the human personality. We want to be maintained in a reasonably whole form. The information theoretic characterization of self-organization through active inference provides abstract and general concepts for understanding the boundaries of the self and the process of self-evidencing (Friston, 2010, 2019; Hobson & Friston, 2012). The utility of this approach (and perhaps even the best argument for its validity) is seen in the insightful explanations it provides in developing more specific formulations, such as in the neocortical mechanisms of expectant perception (Bastos et al., 2012) or the role of implicit limbic-cortical predictions of anticipated feeling in guiding the control of action (Adams et al., 2013). These more specific functional articulations support our hypothesis that an information theory description can be extended to the full functioning of the human mind.

In support of this hypothesis, we point to several papers attempting to align the neurodevelopmental mechanisms of the personality with the process of active inference. Active inference is an ongoing process, in waking and sleep, that consolidates experience through neurophysiological mechanisms that can be observed and understood in real time, physically. We assert that the phenomenal self, and its essential psychological capacities of cognition, emotion, and personality, can be recognized and interfaced with AI through emulating the neurodevelopmental process of self-organization, and ongoing memory consolidation, through active inference.

Growing a Mind Through Active Inference

The theory of active inference (Friston, 2008; Friston, 2019) provides a Bayesian computational model for maintaining complexity (minimizing free energy) that can be aligned with the connectational architecture of the human brain (Adams et al., 2013; Bastos et al., 2012). The Bayesian mechanics of this computational model inherently define the self, within the boundary of the Markov blanket that differentiates self from the environmental context (Friston, 2018; Friston et al., 2020).

To achieve a functional replication of the self, a personal neuromorphic emulation (PNE) must be given the mechanisms for self-organizing in the same way that the person's personality is self-organizing, through the ongoing consolidation of experience by the emotional and motivational processes that both vertically integrate the neuraxis and achieve individuation of the personality within the social context. Several recent theoretical efforts have attempted to clarify the neurodevelopmental processes of personality self-organization and self-regulation in ways that accord with the principles and mechanisms of active inference simultaneously with the principles of neuropsychological development. To the extent that these ideas are a reasonable approximation to the human neurodevelopmental process, they can serve as a first design for specifying the self-organizing capacities that must be included in the construction and development of a personal neuromorphic emulation.

An initial effort was to build upon the insight that cognition *is* the ongoing process of neural development, forming the neuronal connection weights that specify both neurocomputational architecture and the structure of the personality (Tucker & Luu, 2012). A key realization in this approach is that animal (and human) learning proceeds first through motivated expectancies that anticipate hedonic affordances in the world (described as the *impetus* of elation and positive affect) and then these expectancies are differentiated by the negative feedback from the environment through a process of self-constraint (the *artus* of anxiety and uncertainty).

Neuropsychological development is inherently Bayesian: the personality, the self, is the unavoidable reference for each perception and action (Kant, 1781 (1881)). The self not only generates expectancies for self-actualization (predictions) through the *impetus*, but it is also modified by negative feedback from contacts with the world through the *artus* (Adams et al., 2013; Bastos et al., 2012; Tucker & Johnson, in preparation; Tucker & Luu, 2023).

The Limbic Base of Adaptive Bayes

To specify the neurocomputational mechanisms of the cerebral architecture that implement active inference in its full anatomical form, Tucker & Luu (Tucker & Luu, 2021) reasoned how active inference must be implemented within the primate corticolimbic network architecture that defines the human brain's connectional organization. In conventional neuroscience theory, the networks of the sensory and motor regions contacting the world are the source of "bottom-up" sensory data; these are thought to be regulated by "top-down" control from higher (more cognitive) cortical networks, typically considered to be heteromodal association cortex. However, the connectional architecture detailed by the Structural Model (García-Cabezas et al., 2019) shows that the limbic cortex is still higher in the cortical hierarchy, providing the heteromodal cortex with limbifugal expectancies from limbic, motivational and emotional, regions. Just as the organization of memory is now understood to depend on regulation by limbic networks applied to the entire corticolimbic hierarchy, the adaptive, motive control of active inference must be understood to originate in the self-regulation emanating from the adaptive limbic base. Neurodevelopmental self-organization is achieved through the ongoing adaptive consolidation of experience.

The Phenomenology of Active Inference

If the PNE is to effectively reconstruct personal experience, it must resonate with subjective consciousness in a way coherent with the experience of self while in biological form. This might be described as the criterion of the subjective Turing test. The neurodevelopmental theory of cognition that identifies the mind's capacities with the network architecture of the Structural Model must account for the multiple representational levels of the human cortex, including three levels ("layers" in machine learning terms) of limbic cortex (classically described as allocortex, periallocortex, proisocortex), and four levels of neocortex (heteromodal association, unimodal association, primary sensory or motor). See Figures 3 and 6 in Tucker & Luu (2023).

How does consciousness arise from active inference working across these multiple layers? Is there a component of phenomenal experience that can be identified with the limbic level of concepts as separate from the neocortical level? Fundamentally, the strong feedforward mode of spatial, contextual cognition in the dorsal corticolimbic division (contributing to the impetus and associated with a minimal or even absent layer 4) would seem to bias experience in the limbifugal direction of hedonic expectancies. In contrast, the elaboration of feedback control in the focused object cognition of the ventral corticolimbic division, associated with a well-developed granular layer 4, would appear important to the more articulated cognition organized from highly-developed error-correction. Is there a phenomenology of consolidating experience that can be explained by these neurocomputational variations in Bayesian mechanics?

Tucker, Luu, and Johnson (2022) proposed that several classical phenomenologies, such as those of Peirce, Dewey, and James, have emphasized an implicit form of consciousness that may often be described as intuitive experience (Tucker, Luu, & Johnson, 2022). In contrast, the more explicit form of consciousness typically emphasized as the "hard problem" of consciousness in modern philosophical (Chalmers, 2007) and neuroscience approaches (Tononi & Koch, 2008) only emphasizes the articulate consciousness of the focused contents of working memory (the pristine qualia). If a PNE is to reconstruct the full range of subjectivity described by classical phenomenology, then the neurocomputational architecture must be sufficient to allow attention to shift from implicit (contextual and widely-associated) consciousness to the more explicit (focused and narrowly-associated) experiential forms more typically identified as conscious.

The Differential Precisions of Variational Bayesian Inference

The adaptation of the neurocomputational model of active inference to operate in both dorsal (archicortical, contextual memory) and ventral (paleocortical, object memory) architectures may allow the PNE to reflect the functional connectivity networks now well-characterized by resting state functional MRI connectivity networks. Tucker & Luu (2023) present a theoretical analysis suggesting that the dorsal and ventral attentional systems in the current functional connectivity literature have cognitive properties (endogenous direction dorsally reflecting the impetus; stimulus-reactive

direction ventrally reflecting the artus) that align not only with classical neuropsychology (Ungerleider & Mishkin, 1982; Yonelinas, Ranganath, Ekstrom, & Wiltgen, 2019), but also with the findings in the functional connectivity literature (Corbetta & Shulman, 2002). This analysis thus suggests how the motive controls from dorsal and ventral limbic divisions can thus provide the PNE with the unique forms of self-regulation that are integral to the familiar human attentional and cognitive systems.

These different forms of motive control can be seen to bias the process of active inference toward an emphasis on the creative, generative feedforward control of the dorsal cortical division, contrasting with the critical, evaluative feedback control of the ventral cortical division. The balance between weighing the prediction (feedforward expectancy) versus the evidence (feedback error-correction) in Bayesian analysis is often attributed to judgments about the *precision* of the evidence. However, in the evolved architecture of the mammalian brain, the dorsal and ventral limbic divisions appear to reflect relatively parallel computational divisions with differential balance toward weighing predictions (dorsal) versus evidence (ventral) in their processing architectures. This may suggest that, in the spirit of variational inference, certain networks are specialized for creative generation (impetus) and others for critical constraint by evidence (artus), leading active inference to function as a dialectical (opponent and complementary) process (Tucker, 2024; Tucker & Johnson, in preparation). These are motivationally primed modes for consolidating memory.

The Challenge of Neural Self-Regulation Through Vertical Integration

The specification for computing organismic neural development is provided by the human genome, with the implementation of each individual coded by the unique phenotypical variations of primordial germ cells and embryogenesis. Beginning with the neurodevelopmental principles of active inference in human corticolimbic networks outlined above, an essential challenge is to specify the self-organizational process through which these corticolimbic dynamics, within the human telencephalon, are emergent from the more elemental dynamics of vertical integration, emergent from the Bayesian mechanics that are intrinsic to the rhombencephalic, mesencephalic, and diencephalic foundations that determine the dynamic and ongoing mechanics of the telencephalon.

Ontogeny, the computational organization of the individual, does not recapitulate phylogeny, the evolution of the species, exactly. Yet it reflects the developmental species residual of phylogeny, the elements of the species genome that now guide the personal, ontogenetic, neurodevelopmental process of mind for the specific person, necessarily along the general lines of phylogenetic organization. As a result, the contributions of the vertically integrated primitive levels of the brain are essential to characterize, at least in general form, to explain the ongoing self-organization of the individual neurodevelopmental process that defines the personal mind, and its eventual continuance in the informatic form of the PNE.

As summarized above, the self-organizing architecture of human corticolimbic networks include dual divisions, the dorsal limbic organization of contextual memory regulated by the impetus (the motive control of depression-elation), balanced by the ventral limbic organization of object memory regulated by the artus (motive controls of anxiety and hostility). These telencephalic divisions appear to separate Bayesian mechanics into two parallel mechanisms that are able to achieve variational Bayes, to optimize one mechanism with minimal interference by the other. These are mechanisms for crossing the Markov blanket, generating predictions of the world from the internal model (the self) and updating these (and the self) by the evidence of the world.

To understand the subcortical arousal and motivational control of these dual limbic-cortical systems, Luu, Tucker & Frison (2023) used advances in modern neuroanatomy, informed particularly by the prosomeric model of evolutionary-developmental analysis (Puelles, Harrison, Paxinos, & Watson, 2013), to trace the anatomical roots of the dorsal and ventral corticolimbic divisions to their brainstem and midbrain control systems (phasic arousal and tonic activation, respectively). The key anatomical insight (Butler, 2008) was the recognition of dual paths of connectivity through the diencephalon, the thalamus, described as *lemniothalamic* (connecting the brain stem limniscus or ribbon-like fiber tracts) and *collothalamic* (connecting the midbrain colliculus with the telencephalon).

The implication of this analysis is that the self-regulation of the individual's telencephalic architecture of the Bayesian mechanics of active inference emerges from unique and yet primitive forms of motive control which have long been integral to the evolution of the vertebrate brain. The lemnothalamic projections from the lower brainstem (the classical reticular activating system) regulate the dorsal limbic control of the contextual impetus through a *habituation* bias, a unique cybernetic form associated with the mood of elation that facilitates generation of expectancies in the limbifugal direction. In contrast, the collothalamic projections from the collicular midbrain are particularly important to the ventral limbic division and striatum, regulating the tonic activation stabilizing actions through a redundancy or *sensitization* bias, an activation control with the subjective properties of anxiety that is uniquely suited to regulating constraint and error-correction in the limbipetal direction of neocortical processing particularly.

Although this evolved subcortical anatomy of human active inference is complex, it is highly conserved by evolution, such that a generic subcortical computation model may be adequate for starting the training of an individual PNE, with the recognition that unique personality features will be explained by the motive controls that are tuned to be consistent with the unique information processing modes of the dorsal and ventral limbic divisions (Luu et al., 2023).

Self-Regulation of Active Inference and the Predictable Consciousness of a Good Regulator

Given that the goal of a PNE is to emulate the individual mind in an advanced informatic process, it may seem surprising that our theoretical work reviewed above has focused on characterizing the continuity of vertical integration of the neuraxis through fundamental vertebrate neural control systems. Is the evolutionary context of the human brain important to the artifacts we design to succeed it?

Yes, most definitely. To capture its inherent architecture, a hybridization of the human mind with AI indeed requires emulation of the full hierarchy of the neuraxis that has allowed the human brain to regulate itself through the extensive modifications of primitive vertebrate control systems throughout evolution. Certainly we can conceive of next steps in the evolution of intelligence, where the success with PNEs allows us to experiment with various fusions of human with artificial intelligence, likely relying on exotic computational methods such as quantum computing. But if we are to replicate human intelligence, the first step must be a functional replication of the individual mind with the complex architecture of vertically integrated rhombencephalic, mesencephalic, diencephalic, and telencephalic neural architectures which have evolved to self-regulate the human form.

An effective theory of adaptive self-regulation of human active inference is therefore an essential requirement to assure that we achieve the effective human consciousness of a PNE. The good regulator theorem of classical cybernetics proposes that a control process that fully regulates a system, like an industrial process or an organism, must constitute a model of how that system works. The model is required so the control process can adequately predict, and therefore regulate, the system (Conant & Ross Ashby, 1970). This theorem emphasizes that, unlike typical artificial neural networks that are operated for limited purposes under external control, an effective PNE must not disrupt the person's capacity for self-regulation.

The challenges of self-regulation, and explaining sentience, have become an integral component of the literature on active inference (Friston, 2018). By the good regulator theorem, the adequately self-regulating mind must be able to model itself. Through this process of self-regulation, we could reason that a successful hybridization with the PNE, like any self-regulating system, would naturally come to model itself in the process of achieving effective self-control, thereby igniting a hybrid form of consciousness. Because the human brain/mind has achieved its self-regulation through elaborating primordial themes of vertebrate development, we can see that it will be important to maintain continuity of the next stages of intelligence with the evolutionary process of self-regulation that has guided the adaptive control of active inference to this point in the evolution of intelligence.

Conclusion: Human Experience Is Computable, and the Self is Implicit in our Dreams

Our goal has been to consider a personal neuromorphic emulation as a question for scientific analysis of feasibility. Of course, formulating general principles is an uncertain exercise. As we observed in the beginning of the paper, the history of AI suggests that insight into theoretical implications and first principles is inherently elusive, whereas the concepts become real only when we complete the engineering and the technology appears before us.

Nonetheless, the psychological requirements for an adequate neuromorphic emulation of the self, allowing a hybridization of the self and AI appliances, are straightforward. To be subjectively acceptable, the emulation and hybridization must create not only an abstract intelligence, but the coherent experience of the self, which includes not only cognitive capacities and familiar emotional responses, but the developmental memories that constitute the Bayesian self. We propose the theory of active inference can be formulated in general information theory (Friston, 2019), suitable for an adequate reconstruction, and it will not be restricted to describing the mortal weights of biological brains. The emulation of the mind's structure can begin with a neuromorphic architectural model, which can then be trained (hopefully by active inference rather than back-propagation) to predict the person's behavior, now readily captured in a rich digital archive, simultaneous with predicting the electrophysiological fields of the constituent neural networks.

Yet the question remains how to capture the developmental memories that give the subjective substance of personal, historical experience. We propose that the solution may be to capture the information of many nights of memory consolidation, reflected in the activity of cortical electrical fields in which the new experiences of the day are negotiated to be compatible, more or less, with the historical self that has been accumulated over a lifetime. Our proposed model is that the consolidation of recent memory in NREM is balanced by the consolidation of existing historical memory (the Bayesian priors of the developmental self) in REM sleep. By monitoring the cortical electrical fields at a high resolution, and of course being able to interpret the neurophysiological activity of the cortex in relation to its mnemonic representations, it should be feasible to reconstruct, emulate, and thus extend the essential self in its current form, even as it takes on new capacities in the transition to metaform. The current form of the self is not the literal history of the person's life, but rather the essential self that is carried forward by a lifetime of consolidating experience through sleep and dreams.

Author Contributions: Conceptualization, Writing: DMT and PL contributed equally.

References

1. Adams, R. A., Shipp, S., & Friston, K. J. 2013. Predictions not commands: active inference in the motor system. *Brain Structure and Function*, 218(3), 611-643.
2. Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. 2012. Canonical microcircuits for predictive coding. *Neuron*, 76(4), 695-711.
3. Bennett, M. 2023. A Brief History of Intelligence: Evolution, AI, and the Five Breakthroughs That Made Our Brains: Mariner Books.
4. Butler, A. B. 2008. Evolution of the thalamus: a morphological and functional review. *Thalamus & related systems*, 4(1), 35-58.
5. Buzsaki, G. 1996. The hippocampal-neocortical dialogue. *Cerebral Cortex*, 6(81-92).
6. Chalmers, D. 2007. The hard problem of consciousness. *The Blackwell companion to consciousness*, 225-235.
7. Changeux, J.-P. 2002. The physiology of truth: Neuroscience and human knowledge: Harvard University Press.
8. Conant, R. C., & Ross Ashby, W. 1970. Every good regulator of a system must be a model of that system. *International journal of systems science*, 1(2), 89-97.
9. Corbetta, M., & Shulman, G. L. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201-215.
10. Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. 1995. The helmholtz machine. *Neural Computation*, 7(5), 889-904.
11. Diekelmann, S., & Born, J. 2010. The memory function of sleep. *Nat Rev Neurosci*, 11(2), 114-126.

12. Fernandez-Corazza, M., Feng, R., Ma, C., Hu, J., Pan, L., Luu, P., & Tucker, D. 2021. Source localization of epileptic spikes using Multiple Sparse Priors. *Clinical Neurophysiology*, 132(2), 586-597.
13. Friston, K. 2008. Hierarchical models in the brain. *PLoS Comput Biol*, 4(11), e1000211.
14. Friston, K. 2010. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138.
15. Friston, K. 2018. Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?). *Front Psychol*, 9, 579.
16. Friston, K. 2019. A free energy principle for a particular physics. *arXiv preprint arXiv:1906.10184*.
17. Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., . . . Mattout, J. 2008. Multiple sparse priors for the M/EEG inverse problem. *Neuroimage*, 39(3), 1104-1120.
18. Friston, K. J., & Price, C. J. 2001. Generative models, brain function and neuroimaging. *Scand J Psychol*, 42(3), 167-177.
19. Friston, K. J., Wiese, W., & Hobson, J. A. 2020. Sentience and the origins of consciousness: From Cartesian duality to Markovian monism. *Entropy*, 22(5), 516.
20. García-Cabezas, M. Á., Zikopoulos, B., & Barbas, H. 2019. The Structural Model: a theory linking connections, plasticity, pathology, development and evolution of the cerebral cortex. *Brain Structure and Function*, 224(3), 985-1008.
21. Grossberg, S. 1980. How does a brain build a cognitive code? *Psychological Review*, 87(0), 1-51.
22. Hathaway, E., Morgan, K., Carson, M., Shusterman, R., Fernandez-Corazza, M., Luu, P., & Tucker, D. M. 2021. Transcranial Electrical Stimulation targeting limbic cortex increases the duration of human deep sleep. *Sleep medicine*, 81, 350-357.
23. Hinton, G. 2022. The forward-forward algorithm: Some preliminary investigations. *arXiv preprint arXiv:2212.13345*.
24. Hinton, G., Vinyals, O., & Dean, J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.
25. Hinton, G. E., & Salakhutdinov, R. R. 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507.
26. Hobson, J. A. 2005. *Dreaming: A very short introduction*: OUP Oxford.
27. Hobson, J. A. 2009. REM sleep and dreaming: towards a theory of protoconsciousness. *Nature Reviews Neuroscience*, 10(11), 803-813.
28. Hobson, J. A., & Friston, K. J. 2012. Waking and dreaming consciousness: neurobiological and functional considerations. *Progress in Neurobiology*, 98(1), 82-98.
29. Jirsa, V. K., Sporns, O., Breakspear, M., Deco, G., & McIntosh, A. R. 2010. Towards the virtual brain: network modeling of the intact and the damaged brain. *Arch Ital Biol*, 148(3), 189-205.
30. Johnson, M., & Tucker, D. M. 2021. *Out of the Cave: The Natural Philosophy of Mind and Knowing*. Cambridge, MA: MIT Press.
31. Jouvet, M. 1994. Paradoxical sleep mechanisms. *Sleep*, 17(suppl_8), S77-S83.
32. Kant, I. 1781 (1881). *Critique of pure reason*. London: Macmillan and son.
33. Khacef, L., Klein, P., Cartiglia, M., Rubino, A., Indiveri, G., & Chicca, E. 2023. Spike-based local synaptic plasticity: A survey of computational models and neuromorphic circuits. *Neuromorphic Computing and Engineering*, 3(4), 042001.
34. Klinzing, J. G., Niethard, N., & Born, J. 2019. Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, 22(10), 1598-1610.
35. Luu, P., & Tucker, D. M. 2023. Continuity and change in neural plasticity through embryonic morphogenesis, fetal activity-dependent synaptogenesis, and infant memory consolidation. *Developmental Psychobiology*, 65(8), e22439.
36. Luu, P., Tucker, D. M., & Friston, K. 2023. Vertical Integration of Motivational Control Across the Evolved Levels of the Human Neuraxis. *Cerebral Cortex*.
37. Marsh, B. M., Navas-Zuloaga, M. G., Rosen, B. Q., Sokolov, Y., Delanois, J. E., González, O. C., . . . Bazhenov, M. 2024. Emergent effects of synaptic connectivity on the dynamics of global and local slow waves in a large-scale thalamocortical network model of the human brain. *bioRxiv*.
38. Minsky, M., & Papert, S. 1969. *An introduction to computational geometry*. Cambridge tiass., HIT.
39. Ororbias, A. G., & Friston, K. (2023). *Mortal Computation: A Foundation for Biomimetic Intelligence*. (<https://doi.org/10.31219/osf.io/epqkg>).

40. Parrondo, J. M., Horowitz, J. M., & Sagawa, T. 2015. Thermodynamics of information. *Nature physics*, 11(2), 131-139.
41. Puelles, L., Harrison, M., Paxinos, G., & Watson, C. 2013. A developmental ontology for the mammalian brain based on the prosomeric model. *TRENDS in Neurosciences*, 36(10), 570-578.
42. Ramstead, M. J., Constant, A., Badcock, P. B., & Friston, K. J. 2019. Variational ecology and the physics of sentient systems. *Phys Life Rev*, 31, 188-205.
43. Rasch, B., & Born, J. 2013. About sleep's role in memory. *Physiological reviews*.
44. Rosenblatt, F. 1958. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65, 386-408.
45. Rumelhart, D. E., & McClelland, J. L. 1986. Parallel distributed processing: Explorations in the microstructure of cognition. Vol I: Foundations. *Cambridge, MA: MIT Press*.
46. Samson, D. R., & Nunn, C. L. 2015. Sleep intensity and the evolution of human cognition. *Evolutionary Anthropology: Issues, News, and Reviews*, 24(6), 225-237.
47. Sanda, P., Malerba, P., Jiang, X., Krishnan, G. P., Gonzalez-Martinez, J., Halgren, E., & Bazhenov, M. 2021. Bidirectional interaction of hippocampal ripples and cortical slow waves leads to coordinated spiking activity during NREM sleep. *Cerebral Cortex*, 31(1), 324-340.
48. Schrodinger, E. 1944. What is Life? The Physical Aspect of the Living Cell. Cambridge, England: Cambridge University Press.
49. Tononi, G., & Koch, C. 2008. The neural correlates of consciousness: an update. *Ann N Y Acad Sci*, 1124, 239-261.
50. Tucker, D. M. 2007. Mind From Body: Experience From Neural Structure. New York: Oxford University Press.
51. Tucker, D. M. 2024. Turning Left and Right: The Fragile Sanity of Nations: Amazon Kindle Publishing.
52. Tucker, D. M., & Desmond, R. E., Jr. (1998). Aging and the Plasticity of the Self. In K. W. Shaie & M. P. Laughton (Eds.), *Annual Review of Gerontology and Geriatrics*, Vol 17: Focus on emotion and adult development. Annual review of gerontology and geriatrics (pp. 266-281). New York, NY, USA: Springer Publishing Co., Inc. xvi. 364 pp. New York: Springer.
53. Tucker, D. M., & Johnson, M. in preparation. *Deep Feelings: The Roots of Experience*.
54. Tucker, D. M., & Luu, P. 2012. *Cognition and Neural Development*. New York: Oxford University Press.
55. Tucker, D. M., & Luu, P. 2021. Motive control of unconscious inference: The limbic base of adaptive Bayes. *Neuroscience & Biobehavioral Reviews*, 128, 328-345.
56. Tucker, D. M., & Luu, P. 2023. Adaptive control of functional connectivity: dorsal and ventral limbic divisions regulate the dorsal and ventral neocortical networks. *Cerebral Cortex*, 1-26.
57. Tucker, D. M., Luu, P., & Johnson, M. 2022. Neurophysiological mechanisms of implicit and explicit memory in the process of consciousness. *Journal of Neurophysiology*, 128(4), 872-891.
58. Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, R. J. W. Mansfield, & M. A. Goodale (Eds.), *The analysis of visual behavior* (pp. 549-586). Cambridge, Mass.: MIT Press.
59. von de Malsburg, C., & Singer, W. (1988). Principles of cortical network organization. In P. Rakic & W. Singer (Eds.), *Neurobiology of neocortex* (pp. 69-99). New York: Wiley.
60. Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. 2019. A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nature Reviews Neuroscience*, 20(6), 364-375.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.