Temporal Depth in a Coherent Self and in Depersonalization: Theoretical Model

Alexey Tolchinsky,[1] Michael Levin,[2] Chris Fields[3], Lancelot Da Costa[4], Rachael Murphy[5], Daniel Friedman,[6] David Pincus[7]

Abstract

Multiple theoretical models of dissociative experiences have been formulated over the last century. These theories are clinically useful; however, it remains unclear if common factors exist in various pathways leading to an onset of dissociations. This article is the first in a series of papers, where we provide a framework for building an integrated model of dissociative experiences. This framework combines a first-principles-based perspective with dynamical systems perspective. We propose that temporal depth collapse can be a possible common factor in dissociative episodes of any etiology, moreover, we consider this factor to have causal power. In a follow-up paper, we will examine this model from the standpoint of clinical practice and neurobiological theories of dissociations. We will also provide empirical data in support of this model and present ideas for therapeutic applications.

Keywords: dissociation, TAME, temporal depth, self, depersonalization, psychotherapy

[1] Professional Psychology Program. The George Washington University. 1922 F. Street, NW, Washington, DC., 20052, USA, Email: alexeyt@gwu.edu
[2] Allen Discovery Center at Tufts University, 200 Boston Ave. 334 Research East, Medford, MA, 02155, USA
Wyss Institute for Biologically Inspired Engineering at Harvard University, 3 Blackfan St., Boston, MA, 02115, USA
[3] Allen Discovery Center at Tufts University, 200 Boston Ave. 334 Research East, Medford, MA, 02155, USA
[4] VERSES AI Research Lab, Los Angeles, USA
Université de Montreal, Montreal, Canada
[5] Department of Psychiatry at Lehigh Valley Health Network in Bethlehem, Pennsylvania, USA
[6] Active Inference Institute, Davis, CA 95616, USA
[7] University Hospitals of Cleveland, Case Western Reserve University, USA

> There is no self in a given moment: the self is defined by persistence over time.
>
> (Mitchell, 2023)

Introduction

Dissociative disorders (DDs), including dissociative identity disorder (DID) and depersonalization derealization disorder (DPDR), are prevalent in clinical practice. Loewenstein (2018) summarized international epidemiological studies in North America, Europe, the Middle East, and Asia and reported that in clinical samples, including both inpatient and outpatient populations, the prevalence of DDs reached 46%. In a comprehensive review, Boyer (2022) reported that the lifetime prevalence of DDs in the general population was estimated to be higher than that of bipolar disorder or obsessive-compulsive disorder. The International Society for the Study of Trauma and Dissociation (ISSTD), in their third version of the guidelines for the treatment of dissociative disorders, reported that DDs significantly impair patients' functioning and present considerable risk – 67 percent of the patients diagnosed with DDs reported a history of repeated suicide attempts (ISSTD, 2011).

Dissociative disorders are difficult to diagnose. Individuals with DDs, on average, spend from 5 to 12.4 years in some form of mental health treatment before receiving an accurate diagnosis (Boyer et al., 2022). Several reasons have been proposed to account for this, including the clinician's difficulty in imagining this level of psychopathology, the patient's lack of trust in disclosing awareness of their dissociative difficulties and the patient's unawareness that they dissociate. When the diagnosis is reached, outpatient psychotherapy is typically recommended for DDs as the front-line treatment, while pharmacological treatments show marginal efficacy (ISSTD, 2011).

Trauma-related etiological model of DDs appears to have stronger support among clinicians than alternative theories (ISSTD, 2011). More specifically, prolonged elevation of stress accompanied by repeated traumatic experiences in circumstances where a person has no escape (e.g. chronic childhood abuse and neglect) are associated with dissociative conditions (reviewed in Lanius et al., 2018; see also Vonderlin et al, 2018). Clinicians refer to these circumstances as complex post-traumatic stress disorder (C-PTSD, see Herman, 2015), which has a different profile from one or several traumatic exposures, leading to the onset of post-traumatic symptoms (referred to as Acute PTSD). Indeed, approximately 90 percent of individuals with DID in the United States, Canada and Europe experienced childhood abuse and neglect (APA, 2022).

In this paper, we will share a perspective on why the lack of escape from prolonged exposure to unbearable circumstances may lead to the onset of dissociative disorders[1].

This article is the first one in a series of papers that present an integrated theoretical model of dissociations. We hope in this series of papers to highlight one of the common factors that mediates an onset of dissociative symptoms in various etiological scenarios, such as psychological trauma, panic disorder, temporal lobe epilepsy, lesions in the brain, and the use of THC or ketamine. We suggest that despite the important differences in various causes of and presentations of DDs, there is likely a common pathway where various kinds of pathogenesis converge. We hope that this model can help guide future developments in diagnosis and treatment of DDs.

The latest ISSTD (2011) recommendations for psychotherapy of patients with DDs suggest that "treatment should move the patient toward better integrated functioning whenever possible p. 132." We suggest a specific underlying function, which may be necessary for reaching this goal. We refer to this function a "temporal depth" which is indicates how far into the future the agent can plan; this ability necessitates the effective use of the agent's past[2]. A collapse of the temporal depth leads an individual living in the "here and now" accompanied by the inability to access the knowledge of the past or plan for the future. We propose that the restoration of the patient's temporal depth is an essential prerequisite for the stability, coherence, and continuity of the Self.

An additional component of our model as applied to psychotherapy is that both integration and disintegration are necessary at different times during the therapeutic process. We suggest that for the patient who experiences persistent dissociations, some features of DID or DPDR become relatively stable. A shift from these maladaptive regimes toward a more integrated, coherent Self implies a de-stabilization of the attractor landscape corresponding to DID or DPDR and a concurrent stabilization of an alternative 'healthy functioning' attractor landscape. In other words, while the long-term therapy goal should be the improved stability of the integrated, coherent Self, getting there would require a change, which is a destabilization of the maladaptive dynamics.

We find it is useful to contextualize our proposal in the diverse literature on dissociative experiences that accumulated over the course of a century. Ludovic Dugas, who coined the term 'depersonalization' in 1898, was studying the psychopathology of "false memories," including déjà vu (Sierra, 2009). Thus, phenomenology, the patient's subjective experiences was the original method of inquiry. Subsequently, many theoretical models of depersonalization and derealization were developed, including theories implicating the sensory systems, memory, affect, etc. (see Sierra & Berrios 1997 for review).

Some of the current theories of depersonalization and derealization (Deane et al., 2020; Ciaunica et al., 2023a) employ a top-down approach, where these dissociative states were modeled based on first principles, such as the Free Energy Principle (FEP, see Parr et al., 2022 for review). Other researchers chose a bottom-up approach, aiming to find the underlying mechanisms and structures of dissociative experiences (Murphy, 2023; Lanius et al., 2018). In addition, clinicians working for decades with the patients suffering from chronic dissociations share valuable qualitative observations, which add the richness of the patient's subjective data to the abstract theoretical models (Chefetz, 2015).

Such diversity of viewpoints is clearly appropriate for the level of complexity in dissociative experiences. However, one of the challenges related to this multitude of models is that the authors from various disciplines use different terminology and methods of research and no current theory seems to coherently integrate phenomenology, dynamics, neurobiology, and other relevant perspectives. Psychotherapists are often reluctant to read papers with differential equations routinely used in the FEP articles (e.g. Friston et al., 2023a). Similarly, some of the FEP theorists are less familiar with the clinical setting. Clinicians are justifiably concerned when researchers who have no clinical experience opine on how to best help the patients in psychotherapy (Shedler, 2006). Researchers, on the other hand, justifiably state that the qualitative clinical case reports are useful, but often not sufficient to formulate the causal models of the clinical phenomena; and such reports can be augmented with statistics, falsifiable hypotheses, rigorous testing, etc.

We think that all these viewpoints usefully complement each other. Clinicians are correct that the abstract models of dissociative experiences lose the essential qualia. Consider the experience of one of Chefetz's (2015) patients: "At one point I picked up the phone, was talking to my boss [while typing], and saw the words come out of my hands onto the computer screen, but they didn't hit my brain and I had no idea what was going on. (p.125)" Can these subjective experiences be captured in mathematics or neurobiology?

The abstract models of dissociative experiences necessarily coarse grain the subjective human experiences. Such models help us see patterns, differentiate phenomena that are distinct, and make testable predictions. However, this process comes at a cost of losing some of the depth of phenomenology.

An additional issue leading to the possible miscommunications between various theorists and practitioners is the heterogeneity of dissociative experiences. As an example, some clinicians suggest that affective flattening is an essential characteristic feature of dissociative disorders. However, they are not mentioning that post-traumatic flashbacks, which are also a kind of dissociation, are often accompanied by intense feelings, such as helplessness, pain, or rage.

We acknowledge the heterogeneity, which stands in contrast to drawing the bright lines in the definitions of depersonalization and derealization – separating some of them as the "true kind" of dissociations. In agreement with Chefetz (2015), we take an approach of seeing dissociative experiences as heterogeneous and gradual, ranging from common, benign dissociations to more severe, maladaptive forms of dissociations in DID or DPDR.

We hope in this series of papers to provide a possible interface for the collaboration of various disciplines involved and offer a model that is an attempt to integrate these perspectives. This model will necessarily be described in broad strokes as a preliminary framework; we will also discuss the roadmap for the future refinement of it. We will start by describing how we view a coherent and continuous mental/subjective Self from a first-principles-based perspective, including the Technological Approach to Mind Everywhere (TAME, Levin, 2022) and the FEP (Parr et al., 2022). We will then discuss dissociations from the dynamical systems perspective. In the follow-up paper, we will discuss the connections from our model to the contemporary neurobiological and clinical models of depersonalization and derealization.

An important contribution of our model is the highlight the temporal depth collapse playing a role in depersonalization and derealization experiences. A possible relationship between temporal depth and depersonalization has been previously suggested (Deane et al., 2020)[3]. Moreover, Friston (2018) wrote extensively on temporal depth being a necessary component underlying self-consciousness. In our paper, we would like to extend this hypothesis further, to a causal relationship. We suggest that a functional collapse in temporal depth leads to dissociative experiences, including depersonalization.

To clarify, we think that a collapse in a temporal depth can be an intermediate step in the chain of events leading to dissociations; it is unlikely to be an 'original' or the only cause. For example, a problem with the functioning of the person's episodic memory system can lead to the temporal depth collapse and dissociations. However, we claim that a dissociation cannot happen without a temporal depth collapse and a collapse of a temporal depth will reliably lead to a dissociation, making it a necessary and sufficient factor.

Consequently, the therapeutic measures designed to restore and stabilize temporal depth are worthwhile to consider as possible additions to the clinical work with the patients suffering from chronic and persistent dissociations. We will, therefore, highlight the temporal depth collapse as a common pathway in various etiological factors leading to dissociative symptoms[4].

TAME and FEP as Modelling Frameworks

Technological Approach to Minds Everywhere (TAME)

The field of Diverse Intelligence seeks to develop rigorous frameworks for understanding and relating to unconventional minds (Baluška and Levin, 2016; Orive et al., 2020; Pio-Lopez, 2021). This ranges from biologically non-neurotypical humans to the impending plethora of altered, chimeric, and extended beings whose presence will explode outdated binary categories of man vs. machine (Clawson and Levin, 2022; Rouleau and Levin, 2023). One such framework is TAME (Levin, 2022), which is grounded in the biological principles governing the self-assembly of bodies and minds from cells during embryogenesis (Levin, 2019) and the fragmentation of emergent wholes by failure modes such as the morphogenetic dissociation disorder we call cancer (Levin, 2021).

The TAME framework is based on three foundational principles: (a) a commitment to gradualism; (b) an absence of privileged material substrates (material independence); and (c) a commitment to empirical approach to research questions as compared to a philosophical debate in the absence of empirical data. The first principle suggests that there are no bright lines separating various organisms in terms of the complexity of their minds; instead, there is a gradual accumulation of complexity and organization. The second principle suggests that minds are not exclusive to neuron-based systems, or computer hardware-based systems; there is no privileged material that is necessary for the specific kinds of a mind to operate. The third principle suggests that experimental data, rather than opinions or conventions, are the appropriate standard of deciding on how intelligent a system is and how much agency it has.

One of the key concepts of TAME for our paper is the TAME light cone, which is schematically depicted below on **Figure 1**. This concept captures the scale of an agent's ability to use the past experiences to inform its present actions and to plan into the future, as well as the scale of its spatial goals. Put differently, TAME light cone is a measure of the biggest goal that an agent can pursue in space and in time. As you can see in the diagram, a tick operates in its immediate spatial environment and has very limited planning ability or memory of the past. A dog has a larger TAME light cone – it can travel further and can recall and plan more. Humans can support huge cognitive light cones that span the globe and have a time horizon known to be longer than their possible life span.

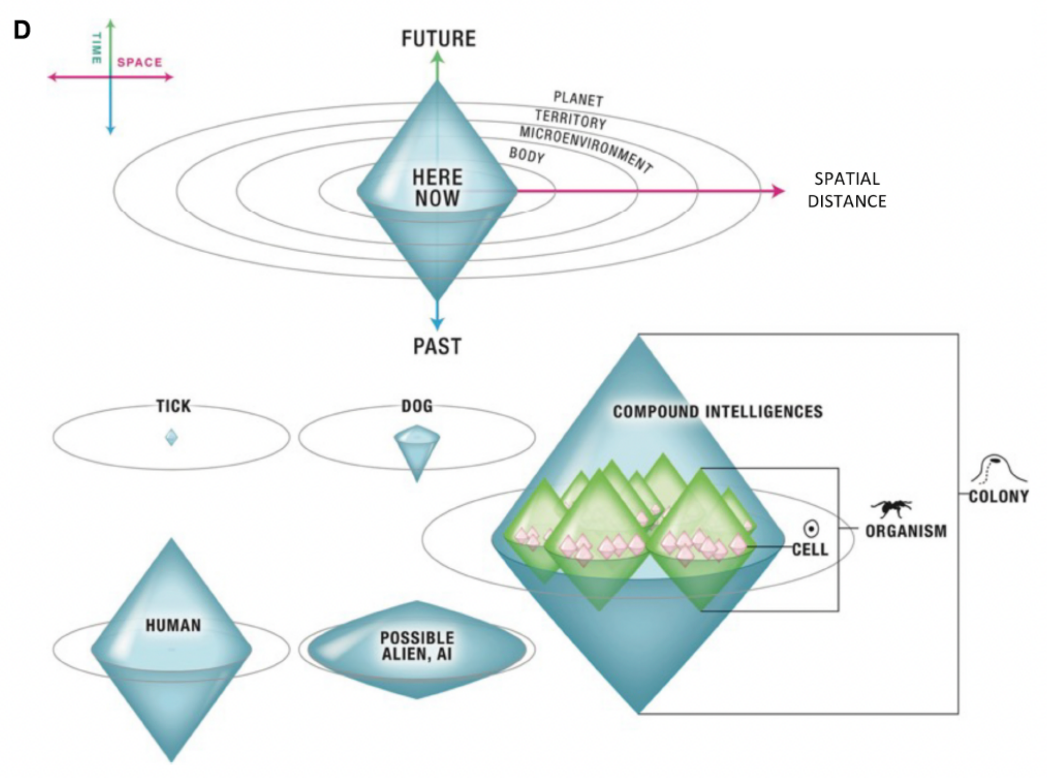

Figure 1. Reprinted with permission from Levin (2022).

The size of the TAME light cone on the vertical axis is related to the focal concept in our paper – temporal depth. You can also see compound intelligences on **Figure** 1, including both a collective of cells - an organism and the collectives of animals, such as an ant colony. Under TAME, all intelligences are collective. An important feature of the TAME light cone concept with respect to the compound intelligences is emergence. An ant cannot build bridges, while an ant colony collectively can accomplish this; and a collective of cells can navigate a maze (Blackiston et al., 2021). What this implies is that the compound Self is not reducible to its components, a whole Self is greater than the sum of its parts, which is another way of saying that the compound Self is a non-linear system (see Rosas et al., 2024 for a technical treatment of the notion of "emergence" implied here).

Levin's (2021) proposed model of a possible etiological pathway to cancer as the result of breakdown in communications between the adjacent cells is, perhaps, one of the most relevant examples of TAME framework applied to the concept of dissociation. Specifically, the closing of the gap junctions of one cell leads to it perceiving the rest of the cells as "not me" or "the environment." This, in turn leads to this newly isolated cell treating the environment as a food source; this cell also reproduces leading to metastasis.

What is essential in this description for our paper is the change in the newly isolated cell's 'identity'. Prior to the closure of the gap junctions and due to the functioning of the intra-cellular communications, the cell collective had a common, cohesive identity, let us label it as "A". The breakdown in the communications between cells effectively led to the fragmentation of A into two parts – the isolated cell, and the collective of cells without it. Should there be a closure of the gap junction in yet another cell in the remaining cell collective, that would in turn lead to further fragmentation into more entities, etc.

When one cell becomes informationally isolated from its neighbors, the previous cell collective's cognitive light cone fragments into several smaller ones, leading to the temporal depth collapse. Therefore, the breakdown in communications between the components leads to both the spatial fragmentation and the loss of temporal continuity.

Free Energy Principle (FEP)

The free energy principle (FEP) was formulated by Karl Friston in the 2000s as a mathematical theory in neurobiology, and extended thereafter to a general theory of living systems (Friston, 2013). One of the key ideas in FEP is that any system that persists will act to maintain its distinction from its environment. Stated more formally, Ramstead (2023) summarized one of the primary FEP claims as follows: “The free energy principle (FEP) says that if the generative model (or dependence structure) of a random dynamical system contains a Markov blanket[5], then it will look as if internal states track the statistics of external states across the boundary.”

The Markov blanket is depicted on **Figure 2**. In a biological organism it is assumed to be composed of Active states and Sensory states.

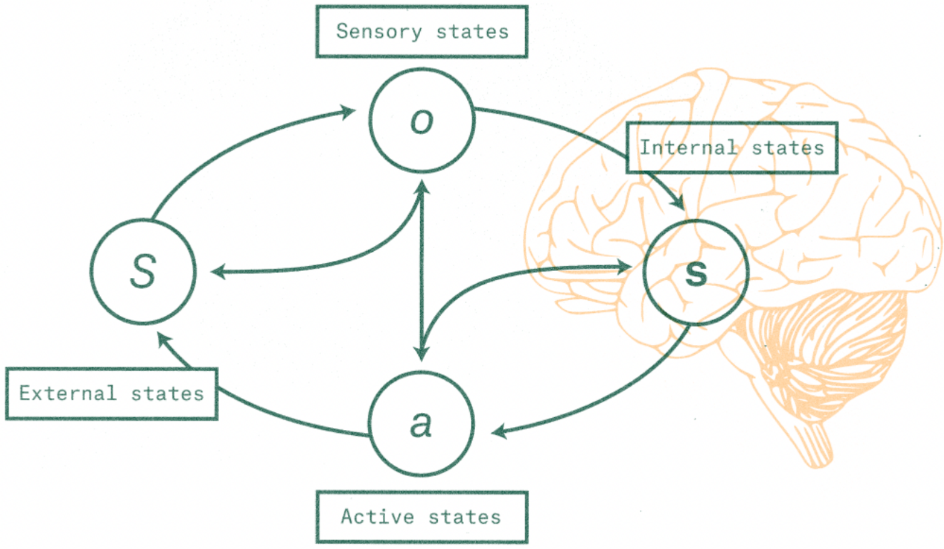

Figure 2. Structure of a Markov blanket as described by the FEP. Formally, a Markov blanket is a set of "boundary" states that separate the "internal" states of some system of interest – here, a brain – from the internal states of its environment. All interactions between the two must pass through, and hence be mediated by, the Markov blanket. Reprinted with permission from Ramstead, 2023

To highlight what is pertinent for our paper, FEP suggests that a system that maintains its existence in the environment necessarily models this environment, consistently with the good regulator theorem (Conant and Ashby, 1970). This model is referred to as "the generative model." The generative model encodes the agent's beliefs about how external states cause its sensory states. For example, if an agent senses warmth, it can infer probabilistically that the sun is shining on it. These models of the world are random dynamical systems that are inferred from and hence adapted to the environment as it changes.

Some systems described in the classical FEP do not have active sates (e.g. a rock). In the presence of active states, the agent is also capable of acting upon the environment, and it does so in such a way to introduce dual environmental changes in order to match the environment to the agent's model (i.e. Active Inference: Parr et al., 2022). For example, a human has an internal belief that she can breathe; should she find herself immersed in water for more than a minute, she will attempt to get back to the surface to reduce the discrepancy between her model of being able to breathe air and the environment.

The generative models under FEP can have variable depth (see figure 12 in Friston et al, 2023b). Some agents are only capable of modelling their immediate environment and live in the "here and now[6]," while others are capable of planning actions into the future. The generative models capable of planning are referred to as "temporally deep generative models." The temporal depth introduced above can be formally described under the FEP as the length of the temporal horizon that is considered during planning, where planning entails counterfactually evaluating the consequences of future courses of action[7] (Friston, 2018).

Table 1. Glossary[8]

Generative Model is a probabilistic model, comprising a likelihood and prior beliefs that specifies how (sensory) consequences are generated by latent (i.e. external) causes, such as hidden states and model parameters.

Inference is the optimization of beliefs by maximizing Bayesian model evidence or minimizing surprise. Approximate Bayesian inference corresponds to minimizing variational free energy.

Active Inference is the minimization of variational free energy through approximate Bayesian inference and active sampling of (sensory) data. This active sampling itself induces posterior beliefs over action, under prior beliefs that action will minimize free energy in the future. This is equivalent to resolving uncertainty with epistemic, information-seeking behavior: see (Friston et al., 2015b; Parr et al, 2022) for details.

Temporal depth (of counterfactual policies) is the depth of temporal processing during planning as inference.

Markov blanket is the set of variables that mediate all (statistical) interactions between a system and its environment. See Parr et al, 2022 for details.

TAME light cone **size** is the size or scale of goals any given system can pursue.

(Variational) Free Energy is a functional of sensory data and posterior beliefs. Free energy scores the surprise of (sensory) data, given posterior beliefs about how they were caused. This furnishes an approximation to Baysian model evidence, aka marginal likelihood.

Interoceptive is pertaining to internal (autonomic) states: see (Craig, 2013; Seth, 2013) for details.

Composite System is a system that consists of multiple components.

Embedded System is a system that is a part of a larger system. Embedded systems interact with their environments, so cannot be considered isolated.

Nested System is a multilayered system with components that in turn contain subcomponents. This process may continue at various scales.

(Fixed) Point Attractor is characterized by the state to which a system evolves over time and to which it returns after being perturbed (Vallacher et al., 2015).

Limit Cycle Attractor is a closed trajectory in the state space that corresponds to sustained oscillations without decay or growth (See Suskick, 2001).

Multistability is a system that "has multiple coexisting attractors and noise is sufficiently strong to cause switching among stable states (Kelso, 2012).

Metastability is a regime where there are no attractors in the system and no energy expenditure is necessary for self-organized tendencies to be visited in turn. Change and persistence are intimately linked in this transient regime. See Tognoli & Kelso (2014) for more information.

A Model of the Subjective (Mental) Self[9]

TAME and FEP perspectives

Material Independent and Belief-based

We see the Self as material-independent and constructed dynamically from an organism's – indeed, any system's – continual efforts at making sense of both its external environment and its internal milieu. Therefore, our model of the Self is not inherently based on the physiology of the human body, including the physiology of the Central Nervous System (CNS). The CNS is just one of the environments where such models can be implemented. As we apply our model to mammals, including humans, we will describe the specific aspect of our model, the Core Self (see ***Hierarchy, Composite System, Boundaries*** below), which is embodied and evolved to help mammals adapt to their habitats and problem-solve in novel environments. While closely related to the body of a specific animal and its natural environment, the specific implementation of the Core Self can also be described in abstract FEP terms, e.g. in Solms (2021).

In addition, we see the organism as a system encoding beliefs (technically, probability distributions) that predict its own states and those of its world. The Self depends upon an organism's ability to infer. The organism's beliefs about its world constitute its inference about the external environment, and the continuum of deficiencies in this process we will call "derealization." The organism's inferences about its own internal milieu – its body and mind – are entailed by the generative model. These beliefs are thought to constitute the subjective experience of the self — we will call this system of beliefs and inferences, as they are experienced by the organism, "the Self;" and the continuum of diverse issues of this process we will call "depersonalization."

Hierarchy, Composite System, Boundaries

As defined above, the Self is a composite, nested, and embedded functional system (please see Compound Intelligences of **Figure 1** as an illustration). We also consider the Self to be a system consisting of hierarchically structured beliefs/inferences, which effectively makes it a dynamic, hierarchical generative model (Parr et al., 2022).

The Self, taken in its entirety, is informationally separated from its external environment by a boundary, a Markov blanket which makes the Self conditionally independent from the non-Self (Parr et al., 2022, p.43). This boundary allows the Self to have a degree of separation and autonomy from its environment. The Self boundary is not material, like skin, but it is an informational boundary through which the "Self" and the "not-Self," its environment, interact. We emphasize that the Self, as we have defined it above, is a model, and is distinct by definition from the system – e.g. the organism – that constructs and implements it. The Self's environment, or not-Self, therefore, includes any components of the organism that are not components of the Self. Therefore, the organism's physiological body is not part of its Self, though the organism's *model* of its body is (in general) part of its Self. The Self is, quite literally, a "construction of the mind."

This is an important point. Any Markov blanket is a boundary in state space, not in physical three-dimensional space. A Markov blanket exists within the causal network of systemic and environmental variables and their causal relationships. While some boundaries happen to be simultaneously informational and spatial (e.g. skin), the Self's Markov blanket is just informational[10]. The Self, being a model, is an informational structure; hence its environment is also an informational structure. The Self's Markov blanket separates, and maintains the independence between, these informational structures.

Under FEP, the Markov blanket is what separates "the thing" from the "not-thing" (Parr et al., 2022). In order for "the thing" to persist in time as a unique entity, the boundary's elements and processes must remain functional and satisfy the properties of a Markov blanket, i.e. it must maintain conditional statistical independence between the "thing" and its environment. The entity informationally demarcated by the Self's

Markov blanket is the Self; the blanket also acts as the interface from the Self to its environment. The Self models its informational environment, it 'senses' it though the sensory states and 'acts upon' it via the active states (Parr et al., 2022).

Various components of the Self are separated from each other by their own functional boundaries, also Markov blankets (see Parr et al., 2022, p.43 for a description of nested Markov blankets). Collectively, all these boundaries play an important role in the stability of the Self and its various components.

As a hierarchical system, the Self has the "Core Self" component at the informational center of the hierarchy, and other components represent more peripheral layers around the Core Self[11]. In our model, the Core Self is the concept that was described by Panksepp and Biven (2012) in Chapter 11 of their book "The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions." In mammals, the essential characteristic of this Core Self is that it is affective. Panksepp and Biven postulated that this Core Self was nonreflexive (anoetic) and dominated by raw affective feelings (primary-process affects in Panksepp's taxonomy) and constituted a part of the purely affective, Core form of consciousness (Solms & Turnbull, 2018).

We take the Solms (2021) view on these raw affective experiences as "felt uncertainty", which is a FEP-based conceptualization. In Solms' model certain organisms do not have affects, but rather inflexible innate reflexes, such as a reflex to approach food and to avoid danger. Affects present an evolutionally advantage to animals that have them. Affects allows an animal to "feel through" the novel problem while using a specific homeostatic mechanism as a guide.

For example, if an animal that had never experienced high heat before were to find itself in a hot place, it could use the internal feeling of the body temperature to guide its actions. The animal will feel better when moving closer to shade and worse when moving away from it. The further the animal's body temperature is from the homeostatic setting point, the worse it feels. A return from the high body temperature to the setting point would be accompanied by a positive feeling of cooling off. When the body temperature returns to the settling point, the feeling of being hot disappears entirely. The system being at or near the settling point suggests that the biological need underlying this affect is met.

This affective mechanism allows an animal to problem solve in novel environments. An organism that has only innate reflexes and no affects is far less likely to survive in completely unexpected circumstances – it would not have an inner "compass" to guide its actions.

A collection of these affective functions that are necessary for the animal's survival constitutes the Core Self. Then, the set of predictions in the Core Self is that all the life-sustaining affects will be at or near their setting points. This state of balance where all biological needs are met corresponds to a minimum in the organism's Variational Free Energy (VFE) – a biologically optimal state. An activation of one of the affects indicates a departure from the VFE minimum, which is a prediction error[12].

Let us now illustrate this concept of the Core Self in neurobiological terms, making concrete some of the abstract terms in the informational model described above. Panksepp and Biven suggested that in mammals, the subcortical structures, including but not limited to the upper brain stem and the periaqueductal grey (PAG), mediated the functionating of the Core Self subsystem. Consequently, the Core Self is thought to be present in decorticated cats and hydranencephalic human children (Solms, 2019) – it does not require a functional neocortex.

The 'higher levels' of the brain's structure in humans, including the neocortex mediate the higher levels of both consciousness and the Self – for example, our abilities to reflect on our own mental states and report them to others, referred to as an 'extended consciousness' (Solms & Turnbull, 2018). Additionally, the neocortex allows humans to have object representations. Then, at the level of the Core Self, we can experience a raw, primitive, wordless, but qualitatively distinct forms of affect, the nonverbal subjective experience: "I feel like this" (e.g. I feel hunger). Solms (2021) suggested that at the level of Core consciousness, without words or images, we can still differentiate a state of hunger from pain – qualitatively and subjectively. To summarize, with the object representation absent, the agent can still experience a raw form of a specific affective distress and then attempt to execute the behaviors to alleviate this distress.

However, with the higher levels of consciousness present, we can bind an objectless feeling to an object, as Solms (2021) describes it: "I feel like this about that (p. 204)." An example of such extension could be "I want an apple." Meta-observations about oneself also rely on object representations. Therefore, observations such as "I look pale" or "I am a pessimist" are various forms of meta-cognition, where the "I" is a recognized mental object being reflected on, described, and thought about.

Let us reiterate this important point, "I" is a meta-cognitive construct, an abstraction, it only exists at the higher levels of the Self (e.g. in Autobiographical Self). It is not used, nor is needed in the Core Self. The Core Self, as a system, has the capacity to detect the affective prediction errors and attempt to minimize the VFE through action without any "I." A meta-cognitive "I" is therefore an illusion in a sense of it not being a concrete object in the world; it is an abstract concept used in language and other forms of meta-cognitive processing (Metzinger, 2004; Seth, 2021; Webb and Graziano, 2015).

One of the reasons Solms and Turnbull (2018) labeled the fundamental, elementary form of consciousness "Core" is because of an asymmetry – the higher levels of consciousness cannot be functional without the Core, while the reverse is not true (Solms, 2021). Solms illustrated this statement with an example from Fischer et al. (2016) that a two-cubic-millimeter size lesion in the parabrachial nucleus reliably induces a coma, while no lesion that size anywhere in the neocortex would cause a cessation of consciousness.

The same can be said about the Core Self with respect to other Self components. Peripheral Self components cannot function without the operational Core Self, which effectively creates a hierarchical and not a heterarchical structure. In addition, as stated in the **Dynamical Systems Perspective on the Health and Pathology of the Self** section, the Core Self has an ability to change the regime of functioning in other Self components by inducing phase transitions. Anatomically, in mammals, this corresponds to the regions of the brain participating in the Core Self functionality influencing the states of the cortical and subcortical brain structures through generalized arousal. We agree with Solms (2021) that the regions in the upper brain stem, including the Reticular Activating System constitute the area upon which consciousness depends; it is the source of arousal and, therefore, of consciousness, without it, no conscious activity (including the Self) is possible.

At the more informationally peripheral levels of the hierarchy, the Self is a composite system containing (a) a Bodily Self (Seth, 2021), an Autobiographical Self[13], a Social Self (Seth, 2021), and other components[14]; and in which (b) each Self component has its own boundary. The Bodily Self refers to a system (generative model) dynamically building inferences about our body, including the various representations and re-representations of the bodily components, interoceptive processing, etc. An Autobiographical Self is a system dynamically representing our life's history. This system relies on both the contextualized event memory (episodic) and the generalized, factual memory (semantic) in humans[15]. A Social Self is a system representing our inferences about how we are seen by others and how we present ourselves and act in the social environment. Each of these components is embedded into the whole Self and it also contains sub-components, creating a nested architecture, as depicted on **Figure 1**.

The non-Core components of the Self are interrelated and influence each other. However, each component can experience a level of dysfunction while the remaining components remain reasonably operational. For example, some level of dysregulation in the Autobiographical Self can be accompanied by an intact functioning of the Bodily Self and vice versa. Thus, a hierarchical, composite structure makes the Self more resilient. With that, a serious dysfunction in the Core Self would lead to a total depersonalization – a complete loss of all aspects of the Self.

If we consider one of the Self's components – The Autobiographical Self, or the Bodily Self, then a coherent and continuous "I am me" also implies that the current instance of the "I" in that subsystem is recognized as matching the representation of "me" encoded in the subsystem-specific memory. Conversely, a prediction error in "I am me" can be seen as an element of depersonalization. While nearly all the low-level components of the underlying physiological architecture (e.g. cells) are replaced throughout the person's lifetime, the continuity of "I am me" is maintained at the level of a belief system, i.e. at the level of the Self as a constructed model.

Each component of the Self has an overall, unified 'identity.' For example, in a Bodily Self it would be 'my body,' which is a belief close to the core of the predictive hierarchy – "my body is coherent, persistent in time and it is all mine". The 'ownership' is an important component of the bodily Self, and the ownership can also experience various forms of dysfunction. Additionally, due to the composite nature of each component, it will contain subcomponents, such as 'my arm.' The prediction errors related to each subcomponent vary in precision, ranging from nonpathological experiences in the rubber hand illusion and escalating to the disturbances that can be seen in somatoparaphrenia or body integrity disorder (BID). All these prediction errors are, among other things, forms of depersonalization. This phenomenon of ‘partial' depersonalization can scale up to an out of body experience (OOB), where the entire body is seen separately as an object.

The Self is Experienced as a Monadic Whole While Being a Form of a Collective Intelligence

According to TAME, all intelligences are collective, while the Self is subjectively experienced by humans as a monadic “whole.” One aspect of this seeming contradiction could be the difference in perspectives – the collective intelligence view is usually the perspective of an outside observer, while the monadic Self is the perspective from within. However, even from this internal perspective, it is not obvious how the coherence of the Self is established and maintained. In our model, the tentative answer to this question is multifactorial, while we realize that it is incomplete.

Seth (2021) shared a viewpoint on a monadic Self as a form of a “delusion” in a sense that it exists only at the level of a subjective belief and not in objective (outside) reality. We agree. Specifically, as applied to humans, we believe that the higher levels of the Self, such as the Autobiographical Self, create an impression of a unified experience, but this is just the experience of the Autobiographical Self and not of the entire system. The Autobiographical Self ‘claims,’ to itself and others, to be the entire Self, while it is not. Then, it is this meta system that is deluded because it “believes its own reports.”

Stated differently, we suggest that the presence of a stable belief "I am whole" in the Autobiographical Self's generative model contributes to us feeling as a monadic Self. Thus, the subjectively perceived unity, the coherence of the Self is an inference.

A second component of the coherence of the Self is related to the informational scale of this phenomenon – the macro scale, as compared to the micro scale of individual neurons or meso-scale of neuronal ensembles. As we move up in the scale of investigation of the brain-mind phenomena, we tend to see the aggregation and coarse-graining of the data. For example, at the macro level of the scalp EEG we lose some data on the variability and noise happening at the micro-level. To illustrate this idea, we can move between the rooms in the house, however, from the standpoint of an observer standing outside, we remain in the same house – there is perception of higher stability/order at the higher scale of observation. This is another pathway of how the Self is experienced as monadic and coherent at the level of the self-conscious, metacognitive mind.

Continuity of the Self in Time; The Assessment of Familiarity/Novelty

Another quality of the Self is its continuity in time. Similarly to the coherence, we suggest that the continuity of the Self in time is an inference. Nearly as a tautology stemming from the definition of a Markov Blanket, the Self will remain "the same" (persist in time) while all the processes/communications across its Markov blanket remain functional.

An additional component of the continuity of the Self and its various components in time is the experience of familiarity, the recognition of the Self to be familiar, not novel. This experience of familiarity can be described as a match, e.g. "I am the same now as I have been in the past."

There is a long history of views on such calculations in neuroscience. As stated in the **Introduction**, Dugas studied deja vu, which can be described as a temporary dysfunction of the 'familiarity functional system,’ where something novel is perceived as familiar, while jamais vu can be seen as a dysfunction in another direction – where something familiar is perceived as novel. We could therefore describe one aspect of depersonalization as being similar to jamais vu – we perceive our body and mind as novel, unfamiliar.

Empirical data support the presence of the ‘familiarity functional systems’ as distinct from other kinds of memory systems; and a version of such familiarity assessment can be present for various mental functions (see Yonelinas et al. 2022 for review). For example, Meyer & Rust (2018) studied the visual recognition in monkeys and have demonstrated that there were dedicated, distributed, dynamical brain-mind systems ('visual recognition memory’) that contributed to the familiarity calculations; and these systems were distinct from other aspects of visual perception. In a different domain of functioning, Darby et al. (2017) suggested that the retrosplenial cortex mediated the calculation of familiarity/novelty as part of the Capgras delusion in human subjects.

It may seem that the familiarity assessment may appear to be similar to a binary on/off switch. However, Rust and colleagues (2018) have shown that the predictive inference framework can indeed be used to perform such calculations and these calculations are probabilistic and not binary. Specifically, in Rust et al.’s view, a prediction error in expecting an object to be familiar constitutes the reaction of novelty; and such prediction errors can happen with variable degrees of precision, they are not all or nothing, even if they appear to us as such. The gradual nature of such calculations allows us to have the reactions such as: “you seem familiar, but I am not sure, did we meet somewhere before?”

To summarize, the familiarity/novelty aspect of depersonalization or derealization can be described by the FEP’s framework of Bayesian inferences, leading to the degree of familiarity being calculated as a continuous variable (Yonelinas et al, 2022). The subjective impression of us dichotomously perceiving something as familiar versus novel can be seen as the coarse graining of such continuous calculations, similar to a categorical perception of having a fever, while the underlying body temperature calculations are continuous.

Representational Capacity

The agent who infers must have a functional representational capacity for representing the world and itself (Da Costa et al., 2021) – a memory system. As noted earlier, the generative model of the Self and its components is hierarchical, or deep (Parr et al., 2022). This implies that the agent is capable of planning into the future, which, in turn, requires an ability to generate, store, and retrieve counterfactual data. When the representational capacity is completely impaired for any reason, the agent loses an ability to infer, leading to both severe depersonalization and derealization. A partial loss of representational capacity, e.g. in some form of amnesia, may lead to some loss of coherence or continuity in either the model of the inner milieu, the environment, or both.

Healthy and Pathological Temporal Depth Changes

Healthy individuals are able to temporarily expand or contract temporal depth voluntarily to some degree through attentional control. Temporal depth is expanded during long-term planning, and is contracted during attention-demanding tasks, e.g.

during “flow” states where successful performance is generally not self-conscious. Involuntary collapses of temporal depth, however, are pathological, and may indicate a dysfunction in an underlying memory system. Should such memory dysfunction happen, an agent who had been capable of recalling the events of the remote past and planning far into the future would lose these abilities. The continuity of the Self, particularly the metacognitive Self, is partially or completely disrupted during such episodes, which is a feature of depersonalization.

It is important to highlight that some degree of depersonalization can be pursued voluntarily, e.g. in meditative practices (Deane et al., 2020) that employ intense concentration to temporarily “suspend” the metacognitive Self, and that when pursued voluntarily, it is not pathological and may even be therapeutic.

Emergence and Non-Linearity

While being a composite, hierarchical system, the Self is not reducible to a set of its components. The functioning of the whole Self is not identical to the functioning of an Autobiographical Self added to Bodily Self and to other components of Selves – these components interact with each other, which creates emergent properties. The Self is therefore a dynamical, non-linear system[16]. We explore this point in more details in the ***Dynamical Systems Perspective on the Health and Pathology of the Self*** section below.

A Relationship Between the Self and Models of the Outside World

As noted earlier, we have a generative model of the external world (the environment) and a generative model of our body and mind, which we call the Self. In some respects, the second one can be considered as operating at a higher level of the predictive hierarchy than the first one. For example, some subcomponents of our Self model may include the inferences about how we infer about the world. An example could be an observation about one's traits, such as "I am a pessimist" – this is an inference about how we model the world.

Several conclusions follow from this observation. One is that these 'meta' parts of the Self tend to operate at the slower time scales than the environment model (Parr et al., 2022). This 'slowing down' of temporal scales is the general trend when we move from the periphery to the center of the predictive hierarchy.

Second, the relationship between the generative model of the environment and of the internal milieu is indeed complex. One can imagine some modelling of the world being functional without any meta-inferences about how this process works – the sentience without an awareness of sentience[17] (Frith, 2021). Indeed, while we are often aware of processing information about the world, e.g. via the feeling of mental effort, we are generally ignorant of how this processing works. And at other times, our observations at the meta-level can lead us to noticing an issue in our interactions with the outside world, such as "I am being distractible."

Together, the meta and the sub models contribute to the hierarchical depth of the generative model and there are many layers of the generative model's hierarchy (e.g. the representations and re-representations of the Bodily Self, according to Craig, 2002).

Graduality

Each Self component has a gradual nature, the degrees of functioning, as opposed to a binary on/off switch for the entire component. What this implies is that depersonalization is a spectrum, and it is heterogenous; it is not a discrete, homogenous phenomenon. For example, some level of dysfunction in the Autobiographical or the Bodily Self can be considered as a degree of depersonalization.

However, the underlying causes of dissociative experiences do not have to be gradual. An acute onset PTSD can lead to the patient developing dissociative symptoms abruptly and unexpectedly. Similarly, an episode of ketamine intoxication, or an epileptic seizure can abruptly result in dissociative experiences.

Model Optimization

Under the FEP, a specific optimization of the generative model takes place – the accuracy is maximized while the complexity is minimized (Parr et al., 2022). What this means for the model of the environment and the Self, is that the size of the cognitive light cone described in the TAME framework does not need to exceed what is necessary for the adaptation of a specific agent to the specific environment. Under FEP, this means that the agent's goals do not exceed the agent's preferred states and the behaviors that contribute to visiting such states. From this perspective, there is a certain economy in modelling. A goldfish needs a larger cognitive light cone than a bacterium (Levin, 2022). Humans are capable of having huge cognitive light cones. Arguably, this allows humans to have the greatest adaptability to the most unusual circumstances for which we have no default (innate) strategies.

With that, most people do not operate in a regime of large light cones most of the time at every scale of the brain-mind functioning. For example, paying attention to some immediate task shrinks the cognitive light cone to the near present. One doesn't daydream while rock climbing, at least not for long. The light cone of an awake and healthy conscious mind may be adaptable to the task at hand. Having a tight temporal focus is not pathological and is sometimes necessary for survival.

Dynamical Systems Perspective on the Health and Pathology of the Self

The generative models corresponding to each component of the Self operate in various regimes in health and pathology depending on the level of generalized arousal and other circumstances (Tolchinsky, 2023). The system's change from one regime to another can be described as a phase transition.

When a healthy human subject is awake, we can describe the state space of each Self subcomponent, such as an Autobiographical Self, as operating in a point attractor regime, corresponding to "I am me." The dissociative experiences in this regime can be mild and benign. The agent returns from these brief fluctuations to the equilibrium point of "I am me." Put differently, if system is mildly disturbed from the equilibrium point, it will reliably return to it; and if it starts from an initial condition of a mild dissociation it will return to the equilibrium as well. Such slight deviations from the point attractor (the lowest plane of the attractor landscape) can be described as operating in the basin of the point attractor (see **Figure 3** for an illustration). The basin consists of all initial conditions that lead to the state of equilibrium "I am me." In such an attractor landscape a dissociation cannot persist, it is only temporary and mild.

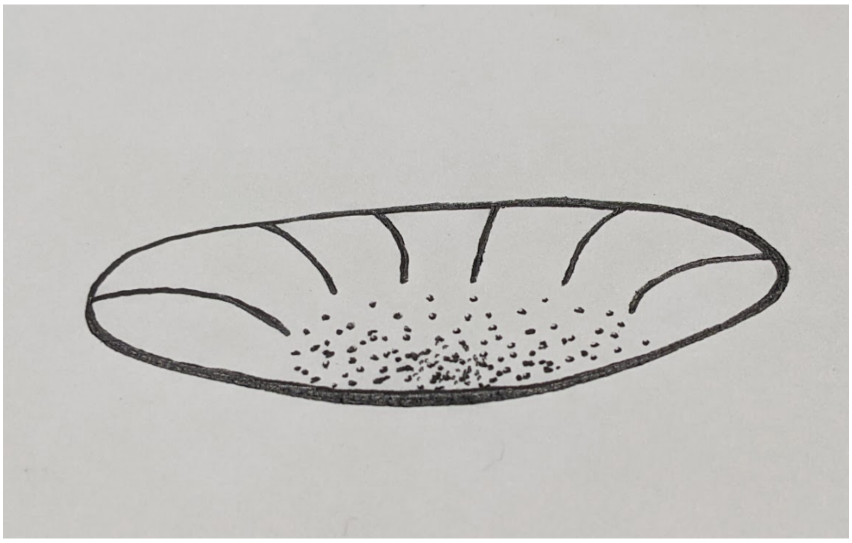

Figure 3. The Self Attractor landscape in health: there is one minimum, corresponding to "I am me," to which paths in the landscape converge.

The lowest point of the point attractor corresponds to a minimum of the VFE. A point attractor regime is stable and without external interference of sufficient power, no change in this regime is expected. An acute psychological trauma is one of the examples of such interferences, which we think can lead to a phase transition, a period of instability, possibly a chaotic regime of functioning. The repeated and lasting traumatization, such as in C-PTSD can also lead to the destabilization of a point attractor regime of the Self. Then, from a temporarily destabilized, possibly chaotic regime of functioning, the attractor landscape can evolve to various new regimes of some stability, corresponding to the specific post-traumatic presentations[18].

One of these presentations is the onset of a disorder with chronic and persistent dissociative experiences such as DID or DPRP. As depersonalizations and derealizations become more intense and frequent, a new attractor/repellor landscape corresponding to these experiences evolves. The onset of DID or DPDR is therefore another phase transition, from a transiently chaotic regime to a landscape where relatively stable states corresponding to dissociations are formed. Thus, when a specific pathological condition "takes root," the system transitions to a multistable mode (Keslo, 2012) with multiple coexisting point attractors. For example, in DID, multiple point attractors emerge corresponding to each of the alters. See **Figure 4** for an illustration.

In DID, when the patient is switching between the alters, we can see an itinerancy, which can be described as a finite set of point attractors, each corresponding to a specific alter in a 'fragmented' Autobiographical Self that has lost its cohesion. The patient's Autobiographical Self being in the state of alter X can be seen as the system moving to the basin of the point attractor "alter X." The repellor regions in the state space between the point attractors can display chaotic dynamics as expected for the boundary region between the two adjacent point attractors. Accordingly, the switch from one alter to another one is not clearly predictable.

Along with the disruption of cohesion of the Autobiographical Self with the onset of DID, the continuity of time is disrupted as well. The switch from alter X to alter Y disrupts the alter X's time continuity. Then, effectively, each alter has its own temporary cognitive light cone that is smaller than the light cone of a coherent Self[19].

Considering the variability in DID, the exact landscape of attractors and repellors may vary depending on the patient's individual circumstances and context. On **Figure 4** you can see that the local minima of the VFE corresponding to each of the alters is surrounded by a global minimum of VFE corresponding to a coherent Self. This is one of the possible options. An alternative possibility, perhaps at a higher level of pathology, is that the VFE landscape has changed so much that the VFE in the minima of the decoherent Self are lower than VFE of coherent one. Such condition can be seen as more stable in psychopathology and therefore more "treatment resistant."

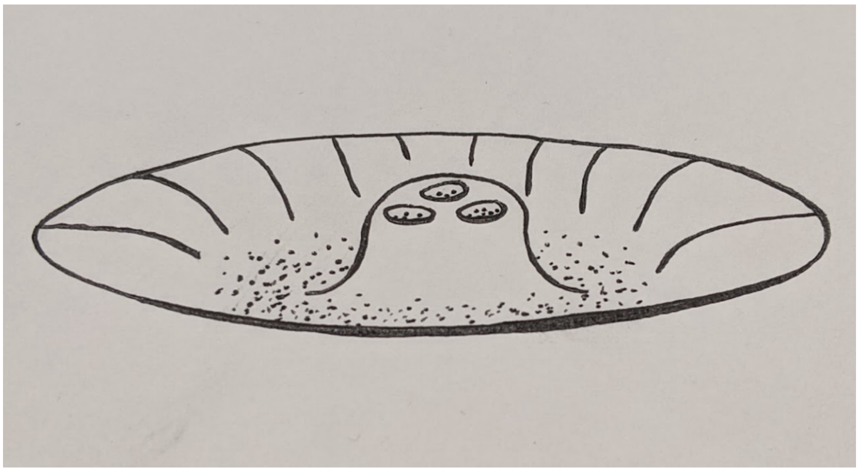

Figure 4. One of the possible attractor landscapes in DID. Here the local minima corresponding to Alters are surrounded by a deeper, global minimum corresponding to a coherent Self.

The attractor landscape corresponding to DID or DPDR may be a relatively stable regime, which is unlikely to change without some form of external interference of sufficient power. Psychotherapy can be such an intervention. In some circumstances, psychotherapy can possibly be augmented with psychedelic treatment or neurostimulation, all of which are various forms of temporarily destabilizing the maladaptive attractor landscape. Then, in treatment, another phase transition can take place to a transiently unstable, possibly chaotic regime with a long-term goal to eventually arrive at a landscape corresponding to healthier functioning[20].

Then, as recovery from DID or DPDR takes place in psychotherapy, the attractor landscape can change back to the single point attractor regime corresponding to a coherent Self, or at least to an attractor landscape that can be seen as somewhat more coherent – with less local point attractors.

While the phase transition from health to DID is influenced by the external factors that are outside the patient's control, such as an exposure to a single or multiple traumatic events, psychotherapy can be seen as a controlled, or guided phase transition.

We can describe the specific processes underlying such phase transitions as follows. The nested and embedded Self, containing the Core Self and peripheral Self components corresponds to a hierarchy of coupled or interconnected attractors. Consistent with FEP, the attractors closer to the Core can be seen as operating at slower time scales. Friston & Keibel (2009) have suggested that a hierarchical system of coupled attractors can be used to describe the phase transitions, such as a slower attractor possibly controlling the phase transitions of a faster attractor. A similar idea called Orbital Decomposition has been proposed by Guastello et al. (1998) for the hierarchical dynamical systems where one chaotic attractor can be decomposed into a series of limit cycle attractors; then, an element of control can be seen in increasing the relative power of one of these limit cycle attractors.

Based on these ideas by Friston and Guastello, in our model, we propose that the attractor landscape corresponding to the Core Self is influencing the phase transitions of the peripheral Self's attractors, such as Autobiographical Self. As stated previously, the Core Self is inherently affective. One of the components of any affective state is generalized arousal. It has been proposed elsewhere that the changes in the generalized arousal level can lead to the phase transitions of the entire neocortex from a periodic to chaotic state and back (Tolchinsky, 2023). Similarly, an acute psychological trauma can be described as an affective 'storm,' starting from the Core Self increasing the level of generalized arousal, leading to higher energy states in the peripheral Self components, which, in turn, may result in the de-stabilization of the 'healthy' point attractor regime in the Autobiographical Self.

An onset of persistent dissociative symptoms in post-trauma can be seen as an adaptation of the peripheral Self components (e.g. Autobiographical Self) into a lower-energy regime, where the affective numbness takes place. This corresponds to the Autobiographical Self "setting down" into a multistable attractor landscape corresponding to DID. Conversely, in an active phase of trauma psychotherapy, the patient is gradually able to tolerate affects to some degree and the Autobiographical Self is moving to a higher energy state, not in an abrupt episode of an affective storm, but in a more gradual fashion. This may be sufficient to cause a controlled de-stabilization of a multistable DID attractor landscape into a temporarily chaotic state, while holding in focus a long-term goal of treatment – to lead the attractor landscape eventually to one corresponding to a coherent Autobiographical Self.

The dynamics described above will influence the temporal depth in the relevant Self components. For example, intact temporal depth is a prerequisite to maintain the temporal continuity in the Autobiographical Self. Then, a transitional, chaotic phase will be accompanied by a temporary collapse in the Autobiographical temporal depth. An onset of persistent dissociations, corresponding to a multistable attractor regime will result in a fragmentation of the temporal depth. To summarize, some stability in the attractor landscape is necessary for the maintenance of a healthy temporal depth in each Self component. Conversely, the phase transitions in the attractor landscape and fragmentations, such as an onset of multistabilty will result in the temporal depth collapse.

Summary

The Self in our model has a nested structure with embedded components and a deep generative model. Stated differently, it is an integrated, multi-layered dynamical system, whose complexity level exceeds that of its individual components. The size of the Self's cognitive light cone is one of the measures of its complexity – its temporal depth.

As mentioned in the ***Introduction*** section, the individuals who experienced prolonged, inescapable exposure to highly stressful environments, accompanied by repeated traumatic experiences, tend to develop dissociative disorders. In our model, we can describe this environmental exposure as lasting stress beyond the agent's ability to manage. Such experience is bound to cause the Self's disintegration – its breakdown into a collection of components, each with a smaller TAME light cone.

This process will be inevitably accompanied by a collapse of the temporal depth. The landscape of the agent's attractors and repellors changes then and the dissociative disorder “takes root.” Then, a sustained effort in psychotherapy is required to help restore Self’s coherence and continuity – the depth of its generative model and its temporal depth.

In contrast to the inescapable, lasting, overwhelming stress, a single episode of drug use can lead to a temporary dissociation due to the transient disruption of the resources necessary to support the depth of the Self’s generative model – its memory systems. For example, an episode of ketamine use may result in the patient’s working memory disruption, leading to a transient dissociation. We may also experience benign daily dissociations on the border of sleep and wakefulness or during meditation. These transient dissociations do not require sustained therapeutic interventions.

Lasting or temporary, severe or mild, dissociative experiences are accompanied by the collapse of the Self’s temporal depth. In this paper, we have shown from multiple perspectives that temporal depth collapse is necessary and sufficient for the onset of dissociations, regardless of the etiology.

In the follow-up paper we will present empirical and clinical data in support of our model and discuss possible therapeutic implications of this model for the patients suffering from dissociative disorders.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Acknowledgments

We are grateful to the reviewers of the paper for their insightful suggestions and to Karl Friston, Evgeny Kaplun, Stephen Guastello, Roxy Wolfe, Richard Chefetz for their valuable comments. Michael Levin gratefully acknowledges support from the

Elisabeth Giauque Trust, London, and Grant 62212 from the John Templeton Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation.

References:

American Psychiatric Association. (2022, October). What Are Dissociative Disorders? https://www.psychiatry.org/patients-families/dissociative-disorders/what-are-dissociative-disorders

Baluška, F., Levin, M., 2016. On Having No Head: Cognition throughout Biological Systems. Front Psychol 7, 902.

Blackiston, D., Lederer, E., Kriegman, S., Garnier, S., Bongard, J., & Levin, M. (2021). A cellular platform for the development of synthetic living machines. Science Robotics, 6(52), eabf1571.

Boyer, S. M., Caplan, J. E., & Edwards, L. K. (2022). Trauma-Related Dissociation and the Dissociative Disorders: Neglected Symptoms with Severe Public Health Consequences. Delaware journal of public health, 8(2), 78.

Ciaunica, A., Levin, M., Rosas, F., & Friston, K. (2023a). Nested Selves: Self-Organisation and Shared Markov Blankets in Prenatal Development in Humans. *PsyArXiv*. https://psyarxiv.com/g8q5d/download?format=pdf

Ciaunica, A., Pienkos, E., Nakul, E., Madeira, L., & Farmer, H. (2023b). Exploration of self-and world-experiences in depersonalization traits. Philosophical Psychology, 36(2), 380-412.

Chefetz, R. A. (2015). *Intensive psychotherapy for persistent dissociative processes: the fear of feeling real* (Norton series on interpersonal neurobiology). WW Norton & Company.

Clawson, W.P., Levin, M., 2022. Endless forms most beautiful 2.0: teleonomy and the bioengineering of chimaeric and synthetic organisms. Biological Journal of the Linnean Society.

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.

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. Nature reviews neuroscience, 3(8), 655-666 https://doi.org/10.1038/nrn894

Craig, A. D. (2013). An interoceptive neuroanatomical perspective on feelings, energy, and effort. Behav. Brain Sci. 36, 685–686. doi: 10.1017/s0140525x13001489

Da Costa, L., Friston, K., Heins, C., & Pavliotis, G. A. (2021). Bayesian mechanics for stationary processes. Proceedings of the Royal Society A, 477(2256), 20210518.

Darby, R. R., Laganiere, S., Pascual-Leone, A., Prasad, S., & Fox, M. D. (2017). Finding the imposter: brain connectivity of lesions causing delusional misidentifications. Brain, 140(2), 497-507.

Deane, G., Miller, M., & Wilkinson, S. (2020). Losing ourselves: active inference, depersonalization, and meditation. *Frontiers in psychology*, 11, 539726.

Donato, F., Alberini, C. M., Amso, D., Dragoi, G., Dranovsky, A., & Newcombe, N. S. (2021). The ontogeny of hippocampus-dependent memories. Journal of Neuroscience, 41(5), 920-926.

Fields, C., Bischof, J., Levin, M., 2020. Morphological Coordination: A Common Ancestral Function Unifying Neural and Non-Neural Signaling. *Physiology* 35, 16-30. http://dx.doi.org/10.1152/physiol.00027.2019

Fields, C., Levin, M., 2020. Scale-Free Biology: Integrating Evolutionary and Developmental Thinking. *BioEssays* 42, e1900228. http://dx.doi.org/10.1002/bies.201900228

Fields, C., Levin, M., 2022. Competency in Navigating Arbitrary Spaces as an Invariant for Analyzing Cognition in Diverse Embodiments. *Entropy* (Basel) 24. http://dx.doi.org/10.3390/e24060819

Fields, C., Levin, M., 2023. Regulative development as a model for origin of life and artificial life studies. *Biosystems* 229, 104927. http://dx.doi.org/10.1016/j.biosystems.2023.104927

Fischer, D., Boes, A., Demertzi, A. et al. (2016), A human brain network derived from coma-causing brainstem lesions. Neurology, 87: 2427–34

Friston, K. J. (2013). Life as we know it. J. R. Soc. Interface 10, 20130475.

Friston, K., Kiebel, S., 2009. Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci. 364, 1211-21

Friston, K., Levin, M., Sengupta, B., Pezzulo, G., 2015a. Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society, Interface* 12. http://dx.doi.org/10.1098/rsif.2014.1383

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G.
2015b. Active inference and epistemic value. Cogn. Neurosci. 6, 187–214.
doi: 10.1080/17588928.2015.1020053

Friston, K. (2018). Am I self-conscious? (Or does self-organization entail self-consciousness?). Frontiers in psychology, 9, 348034.

Friston, K.J., Wiese, W., Hobson, J.A., 2020. Sentience and the Origins of Consciousness: From Cartesian Duality to Markovian Monism. *Entropy* (Basel) 22. http://dx.doi.org/10.3390/e22050516

Friston, K., Da Costa, L., Sajid, N., Heins, C., Ueltzhöffer, K., Pavliotis, G. A., & Parr, T. (2023a). The free energy principle made simpler but not too simple. Physics Reports, 1024, 1-29.

Friston, Karl J., Lancelot Da Costa, Alexander Tschantz, Alex Kiefer, Tommaso Salvatori, Victorita Neacsu, Magnus Koudahl et al. (2023b) "Supervised structure learning." arXiv preprint arXiv:2311.10300.

Frith, C. D. (2021). The neural basis of consciousness. *Psychological medicine*, 51(4), 550-562.

Guastello, S. J., Hyde, T., & Odak, M. (1998). Symbolic dynamic patterns of verbal exchange in a creative problem solving group. Nonlinear Dynamics, Psychology, and Life Sciences, 2, 35-58.

Herman, J. L. (2015). Trauma and recovery: The aftermath of violence--from domestic abuse to political terror. Hachette UK.

International Society for the Study of Trauma and Dissociation. (2011). Guidelines for treating dissociative identity disorder in adults, third revision. Journal of Trauma & Dissociation, 12(2), 115-187.

Kelso, J. S. (2012). Multistability and metastability: understanding dynamic coordination in the brain. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1591), 906-918.

Lanius, R. A., Boyd, J. E., McKinnon, M. C., Nicholson, A. A., Frewen, P., Vermetten, E., Jetly, R., & Spiegel, D. (2018). A Review of the Neurobiological Basis of Trauma-Related Dissociation and Its Relation to Cannabinoid- and Opioid-Mediated Stress Response: a Transdiagnostic, Translational Approach. Current Psychiatry Reports, 20(12). https://doi.org/10.1007/s11920-018-0983-y

Levin, M., 2019. The Computational Boundary of a “Self”: Developmental Bioelectricity Drives Multicellularity and Scale-Free Cognition. *Frontiers in psychology* 10, 2688. http://dx.doi.org/10.3389/fpsyg.2019.02688

Levin, M., 2020. Life, death, and self: Fundamental questions of primitive cognition viewed through the lens of body plasticity and synthetic organisms. Biochemical and Biophysical Research Communications 564, 114-133.

Levin, M., 2021. Bioelectrical approaches to cancer as a problem of the scaling of the cellular self. Prog Biophys Mol Biol 165, 102-113.

Levin, M., 2022. Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds. *Frontiers in Systems Neuroscience* 16, 768201. http://dx.doi.org/10.3389/fnsys.2022.768201

Levin, M., 2023. Collective Intelligence of Morphogenesis as a Teleonomic Process, in: Corning, P.A., Kauffman, S. A., Noble, D., Shapiro, J. A., Vane-Wright, R. I., Pross, A. (Ed.), *Evolution "on Purpose" : Teleonomy in Living Systems*. MIT Press, Cambridge, pp. 175-198. http://dx.doi.org/10.7551/mitpress/14642.001.0001

Loewenstein, R. J. (2018). Dissociation debates: Everything you know is wrong. Dialogues in clinical neuroscience, 20(3), 229-242.

Marks-Tarlow, T. (1999). The self as a dynamical system. Nonlinear dynamics, psychology, and life sciences, 3, 311-345.

Metzinger, T. (2004). Being No One. MIT/Bradford, Cambridge, MA, USA.

Meyer, T., & Rust, N. C. (2018). Single-exposure visual memory judgments are reflected in inferotemporal cortex. elife, 7, e32259.

Mitchell, K. J. (2023). Free Agents. https://doi.org/10.1515/9780691226224

Murphy, R. J. (2023). Depersonalization/derealization disorder and neural correlates of trauma-related pathology: a critical review. Innovations in Clinical Neuroscience, 20(1-3), 53.

Orive, G., Taebnia, N., Dolatshahi-Pirouz, A., 2020. A New Era for Cyborg Science Is Emerging: The Promise of Cyborganic Beings. Adv Healthc Mater 9, e1901023.

Panksepp, J., & Biven, L. (2012). *The archaeology of mind: Neural origins of human emotion.* WW Norton & Company. (No DOI available.)

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: the free energy principle in mind, brain, and behavior*. MIT Press. http://dx.doi.org/10.7551/mitpress/12441.001.0001

Pio-Lopez, L., 2021. The rise of the biocyborg: synthetic biology, artificial chimerism and human enhancement. New Genetics and Society 40, 599-619.

Ramstead, M. (2023) The free energy principle—a precis.

Rosas, F. E., Geiger, B. C., Luppi, A. I., Seth, A. K., Polani, D., Gastpar, M., Mediano, P. A. M. (2024). Software in the natural world: A computational approach to hierarchical emergence. Preprint arxiv:2402.09090v2 [nlin.AO].

Rouleau, N., Levin, M., 2023. The Multiple Realizability of Sentience in Living Systems and Beyond. eNeuro 10.

Seth, A. (2021). *Being you: A new science of consciousness*. Penguin.

Shedler, J. (2006). Why the scientist–practitioner schism won't go away. The General Psychologist, 41(2), 9-10.

Solms, M., & Turnbull, O. (2018). The brain and the inner world: An introduction to the neuroscience of subjective experience. Routledge.

Solms, M. (2019). The hard problem of consciousness and the free energy principle. Frontiers in Psychology, 9, 412177.

Solms, M. (2021). *The hidden spring: A journey to the source of consciousness*. W.W. Norton & Company, Kindle Edition.

Tognoli, E., & Kelso, J. S. (2014). The metastable brain. Neuron, 81(1), 35-48.

Tolchinsky, A. (2014). Acute trauma in adulthood in the context of childhood traumatic experiences. *Neuropsychoanalysis*, 16(2), 129-137.

Tolchinsky, A. (2023). A case for chaos theory inclusion in neuropsychoanalytic modeling. *Neuropsychoanalysis*, 25(1), 43-52.

Vallacher, R. R., Van Geert, P., & Nowak, A. (2015). The intrinsic dynamics of psychological process. Current Directions in Psychological Science, 24(1), 58-64.

Vonderlin, R., Kleindienst, N., Alpers, G. W., Bohus, M., Lyssenko, L., & Schmahl, C. (2018). Dissociation in victims of childhood abuse or neglect: a meta-analytic review. Psychological Medicine, 48(15), 2467–2476. https://doi.org/10.1017/s0033291718000740

Webb, T. W., Graziano, M. S. A. (2015). The attention schema theory: A mechanistic account of subjective awareness. Front. Psychol. 6, 500.

Yonelinas, A. P., Ramey, M. M., Riddell, C., Kahana, M. J., & Wagner, A. D. (2022). Recognition memory: The role of recollection and familiarity. The Oxford handbook of human memory.

[1] In the follow-up paper, we will also discuss dissociations due to non-traumatic etiology, such as those induced by substances or neurological conditions.

[2] See also the formal definition of temporal depth in the Glossary section.

[3] While Deane et al. (2022) suggests a possibility of a relationship between temporal depth and dissociations, we make a claim that a collapse of the temporal depth is necessary and sufficient for the onset of dissociative symptoms. We further suggest that temporal depth collapse can be conceptualized as a common pathway for the onset of dissociative symptoms in circumstances of variable etiology. Finally, we suggest that the focus on restoring temporal depth can be an effective strategy in therapeutic intervensions.

[4] The model of a coherent and continuous Self we present in this paper includes the concepts described elsewhere, e.g. subcomponents of the Self in Chapters 8 and 9 of Seth (2021), a concept of the Core Self in Chapter 11 in Pansepp & Biven (2012), a discussion of temporal depth in Deane et al., (2020). Where our views deviate from these sources, we will articulate the differences as appropriate.

[5] A conceptual boundary between the inside and the outside

[6] Standard artificial neural networks are a good example of this.

[7] It would be reasonable to ask if planning is possible without any memory whatsoever. While a detailed discussion on this topic is beyond the scope of this paper, under FEP, we agree with the following view expressed by Parr et al. (2022): “Forms of planning (and intentionality) can be conceptualized by appealing to the capacity of (some) creatures to select among alternative futures, which in turn requires temporally deep generative models (p.196).” This necessarily implies a functional memory system. Please refer to Parr et al. (2022) for FEP views on the functions of Long-Term Memory (LTM), Working Memory (WM), etc.

[8] These definitions are from Friston, 2018; Levin, 2022; Parr et al. 2022, Kelso, 2012.

[9] We will refer to the mental/subjective Self as simply Self hereafter.

[10] When we use the spatial terms below in this section, such as “center,” “periphery,” “within,” etc., we use them as metaphors.

[11] The Core Self also contains subcomponents and is thus a composite system in and of itself, it is not a monolithic structure. For the sake of clarity, we will not be describing the internal architecture of the Core Self in this paper.

[12] The VFE in this model is dimensional, each affect corresponds to a specific dimension of the VFE. A prediction error results in the organism’s action, which is an attempt to minimize the specific affective dimension of the VFE.

[13] We chose the term Autobiographical Self for what Seth (2021) describes as the Narrative Self due to episodic (autobiographical) memories being present in rats in the absence of verbal narratives.

[14] One of the reasons we do not list each one is that we attempt to describe the overall architecture, including the hierarchy. A second reason is that some of the peripheral components, such as Seth's (2021) "Perspectival Self” are debatable as stable constructs. In agreement with Metzinger, we disagree with Seth on the utility of describing first person or third person perspectives as a specific Self component. A third reason is that the nested, embedded hierarchy means that each these components has sub-components and describing the entire system a list of Selves is not optimal.

[15] Episodic memories are functionally present in mammals, and do not have to be language-based. See Donato et al., (2021) for more details.

[16] This outlook has been proposed before, e.g. (Putnam, 2016; Marks-Tarlow, 1999) and we agree with these authors about the non-linearity of the Self.

[17] We can also add here that the mental states described as “flow” or those non-Egoic Self states achieved by experienced practitioners of meditation could approximate this sentience without awareness of sentience.

[18] The Acute and Complex PTSD are heterogeneous conditions – with immediate or delayed onset and variable constellations of post-traumatic symptoms.

[19] This can be seen as corresponding to Levin's (2021) cancer model of the cells with closed gap junctions – each informationally isolated cell has its own light cone that is smaller than the light cone of the entire cell collective prior to its fragmentation.

[20] The therapeutic interventions can only change the free energy landscape. It is not possible to go uphill an existing, stable VFE landscape. Therefore, in therapeutic work, we may have a goal to erase the current local minima corresponding to the alters in DID or to shift the location of the global minimum.