Article

Not peer-reviewed version

# Causation, Information, and Synergy in the Multiscale Brain Hierarchy

Sergey B. Yurchenko *

Posted Date: 8 September 2025

doi: 10.20944/preprints202509.0678.v1

Keywords: causal chain; partial information decomposition; modular hierarchy; scale transition; multilevel selection

 doi:10.20944/preprints202509.0678.v1



*Article*

# Causation, Information, and Synergy in the Multiscale Brain Hierarchy

**Sergey B. Yurchenko**

Brain and Consciousness Independent Research Center, P.O. 710132, Andijan, Uzbekistan; s.yucko@gmail.com

**Abstract**

The main aim of this paper is to introduce a theory of causation rigorously derived from physical principles and applied to multiscale biological systems. The relationship between causation and information is extensively debated in neuroscience, where causation is involved dually. The first issue concerns how neural activity in the brain causally generates conscious experience. The second issue speculates on how consciousness itself could possess mental power to control the brain. Given the emergent and informational nature of consciousness, mental causation could be admissible under two conditions: downward causation is possible, and information has causal power beyond that provided by matter. Based on the causal set approach in physics, the Causal Equivalence Principle (CEP) allows to evade both of these problems. The CEP is then generalized in terms of the continuity equation in fluid dynamics as the law of conservation of causation, which states that the flow of causation in the universe is conserved across scales. Its corollaries are: (i) there is no preferred scale of observation for causal analysis, (ii) endogenous coarse-graining of biological systems is causally legitimate, and (iii) within the spatial span of the systems, each scale has its own causal structure that cannot be derived from the causal structure at another scale. Which scale tells us the truth about what happens in the universe? The CEP provides an ontological foundation for multilevel selection in evolutionary biology. More broadly, the CEP argues for the stratification of sciences, each operating at its own scale and not reducible to a lower one.

**Keywords:** causal chain; partial information decomposition; modular hierarchy; scale transition; multilevel selection

---

## 1. Introduction

Causal analysis is of great importance in neuroscience and in biology. Downward or top-down causation is a controversial idea, assuming that higher levels of organization can causally influence behavior at lower levels. In neuroscience, downward causation is often related to mental causation or free will, discussed in the context of the mind-body problem. So, Roger Sperry (1980), the Nobel laureate in physiology and medicine for his work with split-brain patients argues:

Conscious phenomena as emergent functional properties of brain processing exert an active control role as causal detents in shaping the flow patterns of cerebral excitation. Once generated from neural events, the higher order mental patterns and programs have their own subjective qualities and progress, operate and interact by their own causal laws and principles which are different from and cannot be reduced to those of neurophysiology.

Can consciousness, as a global product of neural activity, exert causal control over the brain? Or is consciousness a passive emergent phenomenon without causal power? If so, does this mean that consciousness cannot utilize downward causation, or is the concept fundamentally flawed? If the latter is true, what makes this idea so appealing in neuroscience, psychology, and evolutionary biology?

Downward causation is closely linked to various philosophical concepts from complex systems, nonlinear dynamics, and network science, such as emergence, self-organization, and synergy, often

 doi:10.20944/preprints202509.0678.v1



discussed in terms of broken symmetry, criticality, and scale-invariance (Turkheimer et al. 2022; Kesić 2024; Yuan et al. 2024). Synergy is an umbrella term that means an emergent property of complex multiscale systems to spontaneously become self-organized, as encapsulated in the slogan "the whole is greater than the sum of its parts." Examples include flocks of birds, swarms of bees, and ant colonies, which are self-organizing complex systems, demonstrating emergent synergy from interactions of a large number of individual elements (Haken 1983; Kauffman 1993). Downward causation is proposed to explain how a system (the whole) can influence its individual components (parts).

Accordingly, the "strong" version of emergence, associated with downward causation (O'Connor 1994; Bedau 1997), can be linked to higher-order cognitive functions that process the synergistic information and interact by laws and principles that cannot be simply reduced to the underlying neurophysiology (Vohryzek et al. 2022). Regarding the causal role of consciousness, downward causation is implicitly involved in the free will problem, presented in the form of 'synergistic core' (Luppi et al. 2023; Mediano et al. 2022) that could spontaneously emerge in the brain, and govern the underlying neural activity by downward causation over hierarchical levels.

The physics of free will is heavily hinged on the notion of indeterminism. This starts with the question: how could deterministic brain dynamics generate conscious states, which were not predetermined from the past? Different quantum phenomena are suggested as a viable option to account for human (and animal) freedom to decide (Jedlicka 2017; Hunt and Schooler 2019; Yurchenko 2022). In contrast, the neuroscience of free will focuses primarily on Libet-type experiments, which involve comparing two distinctive things: neural activity and subjective experience. Two temporal measures were proposed to compare these variables, known as the readiness potential, detected from the supplementary motor area, and the awareness of wanting to move, reported with the clock. Since the initial experiments conducted by Libet et al.'s (1983), numerous studies have identified a delay between the neural motor predictors and conscious intentions to move around several hundred milliseconds (Schurger et al. 2012; Salvaris and Haggard 2014; Schultze-Kraft et al. 2016). The common conclusion drawn from these experiments is that experiencing free will may be illusory.

A long-standing controversy regarding this conclusion is that readiness potentials and conscious intentions belong to two different and hardly compatible domains: biophysical and psychological (Triggiani et al. 2023). The former operates exclusively on concepts and tools from dynamical models and network science, whereas the latter appeals to subjective and elusive notions such as attention, self-awareness, meta-cognition, and personality. Thus, all these experiments have already been distorted by the mind-brain duality, which has little relevance to the question of whether or not neural activity in the supplementary motor area, or anywhere else in the brain, precedes the emergence of a particular conscious state at a given time. In this sense, if Cartesian dualism is covertly admitted, Libet-type experiments do not threaten the existence of free will at all (Mudrik et al. 2022).

Nonetheless, assuming Cartesian dualism is not sufficient to explain how consciousness might have causal power over the brain. What processes or mechanisms could allow emergent conscious states to influence underlying neural activity? Downward causation has been proposed as a viable candidate for free will. Downward causation is typically characterized as a way, mediated by information flow, that enables a higher level to causally influence a lower level within a system (Farnsworth 2025). In biological sciences, thus, the mind-brain problem acquires a generalized form of neo-Cartesian dualism between information and matter by assuming that in biological (learning) systems information can have causal power beyond that provided by ordinary physical processes. More broadly, it is suggested that the emergence of life may correspond to a physical transition associated with a shift in the causal structure, where information gains direct and context-dependent causal efficacy over the matter in which it is instantiated (Walker and Davies 2013).

Meanwhile, many modern theories of consciousness directly associate consciousness with information that could, in principle, be processed by artificial systems capable of generating machine consciousness (Dehaene et al. 2017). On the other hand, even if the emergence of consciousness is

 doi:10.20944/preprints202509.0678.v1



associated with information processed by the brain (let alone other natural or artificial systems that are not commonly considered conscious), it does not endow consciousness with causal power in the brain. To account for mental causation or free will, these theories implicitly conflate information with causation, and adopt downward causation. In this sense, they can all be classified as theories of strong emergence (Turkheimer et al. 2019). Thus, the age-old problem of free will in the philosophy of mind transforms into the problem of downward causation in neuroscience, which takes the form of neo-Cartesian dualism in biology, where informational terms are all-pervasive (Maynard Smith 2000; Godfrey-Smith 2007).

The purpose of this paper is to disprove downward causation, unless the word "downward" is biased by referring to a spurious, scientifically illegitimate axis in spacetime. The paper is structured as follows: it begins with an examination of causation in various scientific fields, with emphasis on linear causal chains in physics. The relationship between physical causation and information-based measures of causation is then explored. After discussing the concepts of synergy and downward causation, the Causal Equivalence Principle (CEP) is introduced and generalized in terms of the continuity equation as the law of conservation of causation that forbids cross-scale causation in multiscale dynamical systems. The law is then specified in terms of causal scope, scale transitions, and spatial span. Two types of hierarchies (flat and multiscale) are outlined mathematically, showing that information can indeed be synergistic and flow across scales through modular $\subset$-chains but it cannot have causal power beyond that provided by matter. The ensuing discussion shows that the CEP implicitly underlies the renormalization group formalism in physics and provides an ontological foundation for multilevel selection in evolutionary biology. There is no cross-scale causation, but selection operates simultaneously at all spatial scales, each exhibiting its own causal structure, not reducible to a lower one.

## 2. Causation

Causation is a vague notion. In the philosophical literature, it is often suggested to make a distinction between *causation*, defined as the production of one particular event by another, and *causality*, which is regarded as the law-like relation between causes and effects (Hulswit 2002). This distinction is linked to Peirce's view that cause and effect are facts within an epistemological context (in terms of causality), while they are actual events within an ontological context (in terms of causation). In this paper, we will use the word "causation" uniformly to mean the causal analysis of actual events as they unfold in spacetime from the dynamics of physical and biological systems, governed by the laws of nature regardless of observability.

Additionally, the concept of cause is often confused with the concept of reason. A cause is a physical event, associated with the state of a system of interest, which is dynamically followed by another event. Events can be observed at different scales. Causation is evident in the form of canonical cause-effect relationships. Therefore, space, time, and scale are fundamental in understanding causation. In contrast, reason is a cause-like explanation for why something occurred, focusing on logic and neglecting space, time, and scale.

The confusion between cause and reason can be traced back to Aristotle, who defined four classes of causes (*aitia*) that Hofmeyr (2018) called "becauses" or explanatory factors: material, efficient, formal, and final causes, all applied to a thing that should somehow be designed and made of something. Although scientists do not normally think of causation in terms of Aristotelian classification (but see (Ellis 2023)), they still confuse cause with reason, as they are more interested in explaining observable phenomena than in how causation is carried out in time and over spatial scales. For example, a typical formulation in statistics that X (e.g., smoking) can cause Y (e.g., lung cancer) is concerned with a reason, not a cause. Smoking is a bad habit; cancer is a permanent state of health. Neither of these can be regarded as a particular physical event. Another archetypical example of confusing cause with reason is the famous 'chicken-egg problem', where each entity is a reason (not a cause) of the other.

 doi:10.20944/preprints202509.0678.v1



On a strict account, causation should be concerned exclusively with relationships between transient events that can be observed at various spatial scales, such as a sunrise on Earth, a car crash on the road, a fired neuron in the brain, or the detection of a particle in a physical lab.

### *2.1. Physical Causation*

Let us start with the formal definition of an event.

**Definition 1.** An event is an instantaneous state of a system of interest.

Note that the definition does not specify the scale of observation, as the systems of interest can vary in practice. The observation of events depends on two factors: (i) the spatial scale of observation, and (ii) the temporal resolution of observation. Now put aside the traditional view of the world as being full of different things with various physical properties such as position, momentum, size, shape, and so on. Instead, consider linear causal chains that pervade spacetime. From this perspective, there are no things, only instantaneous events that represent their dynamics.

Causal analysis becomes much more rigorous when causation is represented by linear causal chains, as conceptualized by the Causal Set Approach based on Lorentzian geometries of spacetime (Bombelli et al. 1987). This approach follows the principles of relativity theory, which specify that the speed of causal action cannot be faster than the speed of light. The finite speed of causation entails three consequences: (i) the cause must necessarily precede the effect, (ii) simultaneous events are mutually causally independent within a fixed frame of observation, and (iii) linear causal chains must satisfy the Markov property.

The linearity here means that any causal chain evolves only at the same scale and can be graphically depicted as a worldline in Minkowski space $(M, g)$. Formally, if linear causal chains are defined on a vector space $V$ where the link between two nearest events is a vector symbolizing their cause-effect relation, then the linearity is defined traditionally via a linear map $V \to W$ preserving additivity and homogeneity. These yield the so-called superposition property, which states that the effect caused by two or more events is the sum of the effects caused by each event separately. An immediate consequence of this is that every macroevent, observed at the macroscale, can be decomposed into the "sum" of simultaneous and, therefore, mutually causally independent microevents at the microscale.

A causal set is presented by a partially ordered set $\mathcal{L} = (M, \prec)$, with a binary relation $\prec$, which symbolizes causal order in physical spacetime and corresponds in relativity theory to a timelike interval between two events in Minkowski space $(M, g)$ (Sorkin 2009).

**Definition 2.** A causal set $\mathcal{L}$ is a set of elements (events) that satisfies the following conditions:

$$\text{irreflexivity: } (\forall x \in M)(x \not\prec x); \tag{1.1}$$

$$\text{transitivity: } (\forall x, y, z \in M)\, x \prec y \,\&\, y \prec z \Rightarrow x \prec z. \tag{1.2}$$

Condition (1.1) states that no event can be a cause of itself. Together with condition (1.2) they imply that linear causal chains in $\mathcal{L}$ cannot contain closed loops. Intuitively, this follows from the uniqueness of events in spacetime. We can observe the same event repeatedly, but each occurrence is unique in time. If a unique event $x$ causes two independent (simultaneous and unique) events $y$ and $z$, then the linear chain, containing $x$, splits into two linear chains, one containing $y$ and the other containing $z$. Conversely, two linear chains, containing events $x$ and $y$ separately, converge at event $z$ if $z$ is caused by both $x$ and $y$.

The causal chains can be divided into "one-body" linear chains concerned with only one body (e.g., a simple pendulum) and "multi-body" linear chains involving many bodies (e.g., Newton's cradle). In the former case, a linear causal chain can be described as a Markovian process by the temporal evolution of a system whose instantaneous states are events, each causally dependent on the previous one. In the latter case, when two (or more) systems interact, the resulting state of each of them is caused by both its own previous state and the state of the other system before the interaction. For example, a collision between two solid bodies, each with its own causal history, is an

 doi:10.20944/preprints202509.0678.v1



event that impacts the subsequent states of both bodies (including their energy and momentum in spacetime). Thus, linear one-body causal chains can converge and split at different events, generating multi-body linear causal chains. Both types of chains pervade spacetime as the global causal set $\mathcal{L}$. Furthermore, since events are observable at various scales and due to the linearity of causal chains, $\mathcal{L}$ is not confined to a single preferred scale but should pervade spacetime at all scales.

Consider the murmuration of starlings, often presented as an example of emergent synergy. The flock of starlings contracts, expands, and even splits, continuously changing its density and structure as if it has a 'life of its own,' distinct from the thousands of individual birds which constitute the flock. Numerical 3D-simulations of a flock demonstrate that each bird should interact on average with a fixed group of neighbors from six to seven by relatively simple rules to exhibit typical emergent phenomena (Ballerini et al. 2008). In dynamics, the instantaneous states of birds are the microevents that causally impact each other, impelling neighbors to change their flight path in response to their actions. The feedback, repeated over time, generates reciprocal causal loops which are not, however, temporally closed in $\mathcal{L}$. Instead, there is a set of entangled linear multi-body causal chains at the scale of individual birds, involving avalanches across scales (Cavagna et al. 2010) and other features, indicative of self-organized criticality on the edge of chaos and order (Adami 1995). These macroscopic features manifest themselves only at the scale of the flock, whose instantaneous states we observe as macroevents that unfold in spacetime as a linear one-body causal chain from which synergistic phenomena emerge. Thus, one-body and multi-body causal chains not only can co-exist within a dynamical system but also pervade it at different scales.

### *2.2. Causal Reasoning in Statistics*

The causal set $\mathcal{L}$ can be locally represented by a directed acyclic graph $G = (N, E)$, where $N$ is a set of nodes, with $E \subseteq N \times N$ being the set of edges between nodes. Intuitively, if nodes in $G$ are associated with physical events, Bayesian networks for counterfactual causal modeling can then be imposed upon the linear causal chains in $\mathcal{L}$ by ascribing random variables to nodes, with edges representing the conditional probability for the variables (Pearl 2000). This makes it possible to use "causal loops" in data analysis, where nodes are associated not with actual physical events but with phenomena (e.g., homeostasis) or categorial abstractions (e.g., age-Alzheimer's disease) taken as the variables of the graph to detect statistical dependencies between them. In fact, Pearlian causal modeling is more concerned with reasonable explanations of regularities than with actual causal chains as they unfold in spacetime on their own by the laws of nature regardless of whether or not we can observe them. While the passage of time is coarsely grained in data obtaining, the scale and causal order are generally neglected in the probabilistic analysis of the data. Thus, there is a principled paradigm shift from explaining actual observer-independent linear causal chains to obtaining reasonable explanations of observable dependencies (Woodward 2003).

In neuroscience, an injection of propofol or the administration of a drug at a molecular scale are said to cause loss of consciousness or promote mental health respectively, both defined as examples of upward causation. In the same mode of counterfactual reasoning, a high body temperature, which is a weakly emergent property of a system, resulting from Brownian fluctuations of cell components in an organism, can be called a cause of mortality among patients. Although these examples propose verifiable cause-like explanations, they abandon the domain of physical causation, applicable exclusively to transient states of a system of interest at different moments of time, depending on a temporal resolution provided by observation. What is important here is that confusing cause with reason can make downward causation admissible as well, e.g., by saying that the environment exerts large-scale constraints on organisms through downward causation (Noble et al. 2019; Ellis and Kopel 2019). Somewhat ironically, counterfactual reasoning allows to turn the above examples of upward causation into downward causation by merely shifting the scale of observation in the so-called "fat-handed" interventions (Romero 2015), e.g., if an injection of propofol and the administration of a drug are defined as environmental constraints, imposed upon the patient's brain by a clinician in a lab.

 doi:10.20944/preprints202509.0678.v1



*2.3. Causation and Prediction*

We can observe a similar shift from actual causation to cause-like explanations in most famous causation measures such as Granger causality or Transfer entropy, which are formulated in terms of predictive power. Clearly, if it can be predicted that the occurrence of event X always entails the occurrence of event Y, then there is likely a linear multi-body causal chain between them. However, drawing this conclusion in the context of reason, may confuse or even ignore the scales of description between the variables of interest. Although these measures respect causal order, fine temporal and spatial resolution is limited in practical applications. Coarse-grained causal modeling is scientifically legitimate when applied correctly, but mixing different scales can create a loophole for downward causation. Some proponents of strong emergence directly equate coarse-graining with downward causation (Hoel 2017; Grasso et al. 2021). The following section will explain how this scientific bias arises from conflating causation and information in the context of linear causal chains.

## 3. Information and Causation

What makes information theory a useful analytical tool in neuroscience is its model independence, which is applicable to any mixture of multivariate data, with linear and non-linear processes (Wibral et al. 2015; Timme and Lapish 2018; Piasini and Panzeri 2019). However, its applicability to causal analysis must be taken with caution. Information theory was originally developed by Shannon (1948) for the reliable transmission of a signal from a source to a receiver over noisy communication channels. It was demonstrated that the maximal capacity $C$ of a discrete memoryless channel with input $X$ and output $Y$ is given by mutual information:

$$C = I(X;Y) = H(X) - H(X|Y), \quad (2)$$

where $H = -\sum_{i=1}^{N} p(x_i) \log p(x_i)$ is Shannon entropy.

Since the physical nature of a signal, the length of a channel, and time for transmitting information are not conditioned, coarse-graining is explicitly embedded in the definition of entropy: $H$ is independent of the nature of signal and of how the process of transmitting is divided into parts, or, in Shannon's (1948) own words: "if a choice be broken down into two successive choices, the original $H$ should be the weighted sum of the individual values of $H$." More formally, Shannon entropy is an additive measure: $H(X,Y) = H(X) + H(Y)$.

*3.1. Information-Based Measures of Causation in Neuroscience*

How can all of these elements, concepts, and information-theoretic measures be interpreted in the causal analysis of neural networks? First, the channel can be conceptualized as a tube that is maximally isolated from the environment for transmitting a linear one-body causal chain at an appropriate spatial scale. Reducing temporal resolution also allows us to "compress" the causal chain into a single pair, where the cause is an input $X$ and the effect is an output $Y$, omitting all intermediate events ("choices") between them. Second, in neural networks, neurons can take the place of both the source and the receiver, while their instantaneous states represent events. Accordingly, synaptic connections provide the communication channels for a single causal pair between two neurons, which are the input $X$ and the output $Y$ respectively (Figure 1a).

In neuroscientific studies, the temporal and spatial resolution, provided by various neuroimaging techniques, is reduced by ascribing input/output locations not to single neurons but rather to brain regions. Mutual information $I(X;Y)$ is a coarse-grained measure that tells us how much our ignorance about one part of a system is reduced by knowing something about a different part of the system. Its value is zero when $X$ and $Y$ are causally independent (there is no synaptic link between them) so that observing one tells us nothing about the other. Mutual information is symmetrical and upper-bounded by Shannon entropy:

$$I(X;Y) = I(Y;X), 0 \leq I(X;Y) \leq H(X). \quad (3)$$

The symmetry property makes this measure less appropriate for causal analysis since it does not discriminate the causal direction from an input (source) to an output (receiver). Although in

 doi:10.20944/preprints202509.0678.v1



engineering communications, the source and receiver are known so that the causal order between events is naturally preserved, one of the main goals of causal analysis in neuroscience is to unravel the causal structure of fine-grained synaptic circuitry in the brain. Contextually, mutual information is a measure of functional coarse-grained connectivity between large-scale neural networks, derived from statistical correlations.

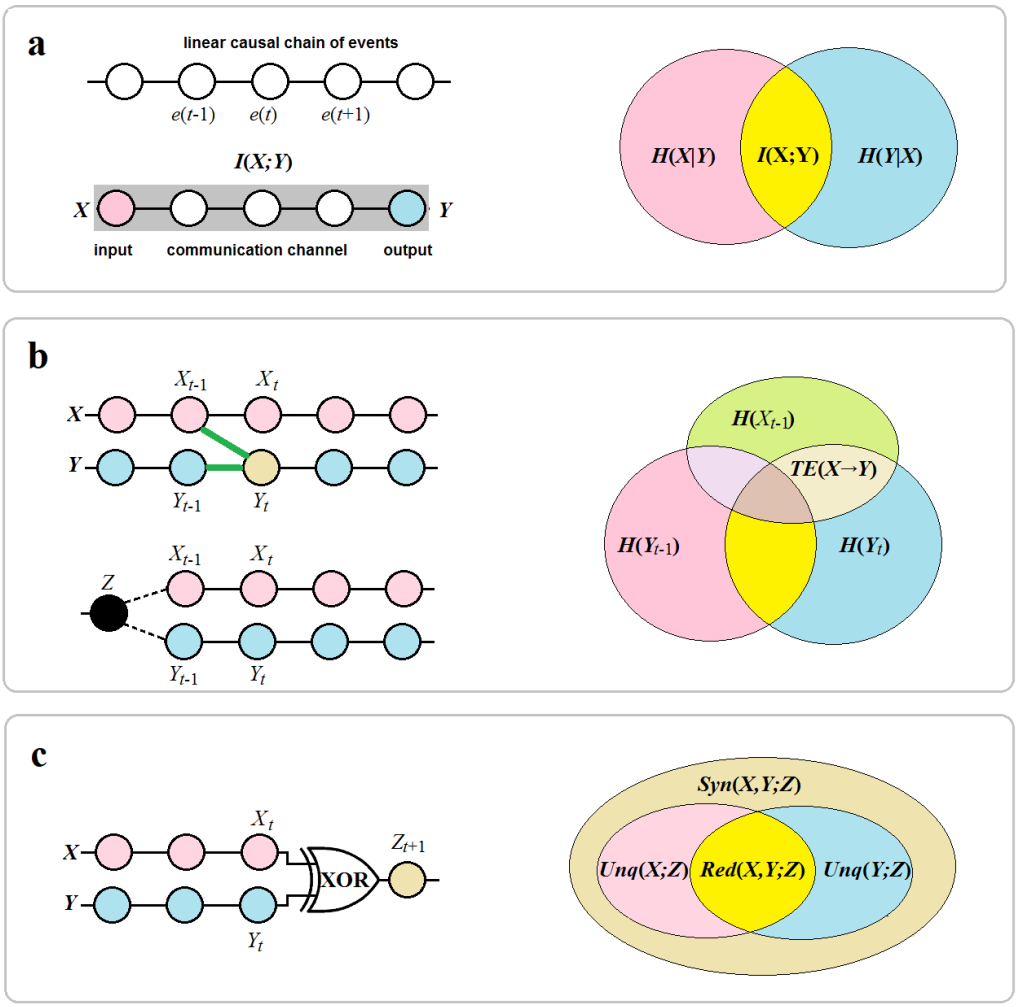


**Figure 1.** (a) In engineering, communication channels act as physical transporters of linear causal chains. Examples of these channels in biology include blood vessels or white-matter fibers. Mutual information can be applied to these causal chains. (b) Transfer entropy between two multi-body (top) or one-body (bottom) causal chains allows to statistically measure the presence of a causal link. (c) Partial information decomposition (PID) allows to decompose mutual information two (or more) sources provide about a target into redundant, unique and synergistic components, making the whole greater than its parts.

In contrast, a more advanced measure, known as transfer entropy, can detect effective connectivity in a verifiable manner (Ursino et al. 2020). Transfer entropy (TE) is a measure of directed information transfer between two (or more) processes in terms of predictive information by observing how uncertainty on the present measurement of $Y_t$ is reduced if knowledge of the past of $X_{t-1}$ is added to knowledge of the past of $Y_{t-1}$:

$$TE(X \to Y) = I(Y_t; X_{t-1}|Y_{t-1})$$
$$= H(Y_t|Y_{t-1}) - H(Y_t|Y_{t-1}, X_{t-1}). \quad (4)$$

 doi:10.20944/preprints202509.0678.v1



TE is asymmetric and upper-bounded by mutual information (Figure 1b). This is then expressed in terms of causation: if a signal A has a causal influence on a signal B, then the probability of B, conditioned on its past, is different from the probability of B, conditioned on both its past and the past of A, which shows a close analogy to Granger causality (Barnett et al. 2009). Since deriving a causal structure from complex systems such as the brain is challenging, many studies suggest that estimating directed information through TE can be an effective diagnostic tool for inferring causal relationships (Wibral et al. 2015). TE can capture causal order but only by virtue of preserving temporal order. There is evidence that this measure can sometimes fail to detect a causal link when it exists, and sometimes can suggest a spurious link (Lizier and Prokopenko 2010; James et al. 2016; Tehrani-Saleh and Adami 2018). In fact, TE measures correlations that can result from a direct causal effect via an edge between two nodes $X$ and $Y$ in neural networks, indicating a causal link between two events in two separate linear causal chains in the brain. On the other hand, long-range correlations can also appear due to a common cause $Z$ of events in the past without a causal link between them (Figure 1b).

### *3.2. Synergistic Information from Multiple Resources*

Since Shannon entropy is additive, mutual information underestimates the synergistic properties of information that can emerge from multiple inputs, such as stereoscopic vision in 3D space provided by the two eyes. More generally, a system exhibits synergistic phenomena, if some information about the target variable $Z$ is disclosed by the joint state of two (or more) source variables that is not disclosed by any individual variable $X$ or $Y$. Williams and Beer (2010) had proposed Partial Information Decomposition (PID) which allows for the division of $I(X,Y;Z)$ into information "atoms" as follows:

$$I(X,Y;Z) = Red(X,Y;Z) + Unq(X;Z) + Unq(Y;Z) + Syn(X,Y;Z), \quad (5)$$

where $Red(X,Y;Z)$ represents the redundant information about $Z$ contained in both $X$ and $Y$, $Unq(X;Z)$ and $Unq(Y;Z)$ correspond to the unique information provided by $X$ and $Y$ separately, and $Syn(X,Y;Z)$ refers to the synergistic information that can be derived from $X$ and $Y$ together but not from each of them alone.

The simplest example of a synergistic network in engineering is one in which $X$ and $Y$ are independent binary variables, and $Z$ is determined by the XOR function, $Z = X \oplus Y$ (where $\oplus$ is the XOR operator). It can be shown that the mutual information between the source variables and target variable vanishes, $I(X;Z) = I(Y;Z) = 0$, which implies that neither of them alone provide information about $Z$. However, together they completely determine its state. The relationship between $Z$ with $X$ and $Y$ is called "pure synergy" since the value of $Z$ can be computed only when both $X$ and $Y$ are known (Figure 1c). Although this technical example helps our intuition, it does not capture the essence of synergy as extra information (non-additive bonus) beyond the information, provided by the sources separately. Especially, the XOR gate has nothing to do with the large-scale patterns of synergy, such as the spontaneous self-organization of complex systems in the absence of external guidance (Haken and Portugali 2016). Instead, this example shows how transfer entropy can be blind to direct causal links. Because mutual information between source variables and target variable in XOR networks is zero, transfer entropy vanishes too, $TE(X \to Z) = TE(Y \to Z) = 0$, despite the obvious fact that they both causally affect the state of $Z$.

Synergy could be better described in the cryptographic context, where access to a secret is distributed among the participants, each of which holds some unique information about the secret. Thus, it should explain how a synergistic (non-additive) component, provided jointly by two or more sources, can make mutual information greater than the sum of the individual information contributions provided by the sources (Gutknecht et al. 2021):

$$I(X,Y;Z) \geq I(X;Z) + I(Y;Z). \quad (6)$$

In particular, since synergistic information is inherently non-additive, its PID-representation via Venn diagram is apparently inconsistent in set-theoretic terms: the whole $I(X,Y;Z)$ is greater than

 doi:10.20944/preprints202509.0678.v1



the sum of its parts $I(X;Z)$ and $I(Y;Z)$ as if $Syn(X,Y;Z)$ would arise *ex nihilo* like magic (Figure 1c). An explanation comes from the fact that unlike the other three "atoms" in Equation (5), the synergy component emerges at a larger scale on the network of sources. Indeed, redundant information can be completely retrieved from any one source at a corresponding scale. Unique information is also provided at the same scale by each source alone. In contrast, synergistic information can only be retrieved from all the sources at a larger scale, corresponding to their union. Missing even one portion of that extra-information from a single source could destroy synergistic information.

To generalize aforesaid, consider a dynamical system $\boldsymbol{X_t} = (X_t^1, \dots, X_t^n)$ of $n$ variables, evolving over a discrete (Markovian) stochastic process by a time lag 1 in a state-space $\mathcal{X}$. At first sight, the amount of synergistic information, provided by the system $\boldsymbol{X_t}$ should be represented by the sum of the portions within $I(X_t^i; Z_t)$ that are provided by each source $X^i$ about $Z$ at time $t$. On the other hand, each portion of this extra information should initially reside within $Unq(X_t^i; Z_t)$ at the scale of each $X^i$ but emerge only at the scale of their union at time $t+1$. As stated, Shannon mutual information does not respect the causal order between inputs and outputs (unless explicitly given). It also does not discriminate between scales. Nonetheless, temporal order can still be imposed on information-theoretic measures such as time-delayed mutual information or transfer entropy by taking into account that transmitting information across scales requires time.

Let us return to stereoscopic vision, and perform a simple thought experiment. If we close one eye, we will only receive mutual information between a target and a source (the second eye), losing all PID-components in Equation (5). Now if we have both eyes open at time $t$, do we (our brains) receive synergistic and unique information simultaneously or does the synergistic (stereoscopic) effect occur shortly after? If the latter is true, we could explain Equation (6) (and the set-theoretic inconsistency of synergy in Figure 1c) via time-delayed mutual information as follows:

$$Syn(\boldsymbol{X}_{t+1}; Z_{t+1}) = \sum_{i=1}^{n} Unq(X_t^i; Z_{t+1}). \qquad (7)$$

In this sense, Equation (5), presented in a timeless form, is not entirely correct: the components $Red(X,Y;Z)$ and $Syn(X,Y;Z)$ do not occur simultaneously but decompose mutual information $I(X,Y;Z)$ by a time lag 1. Informally, synergistic information is mutual information that emerges at the macroscale of a system from its unique information components provided by the system at the microscale, but with a time delay. This makes synergy a function of both time and scale. We can now interpret this component in the context of Equation (7). To bolster intuition, consider a thermodynamic process of the growth of a crystal in a supersaturated metastable solvent. In this physics-inspired scenario, the redundant information $Red(\boldsymbol{X}_t; Z_t)$ is like a seed crystal within the solvent. This seed is necessary for triggering the growth of a crystal lattice, i.e., $Syn(\boldsymbol{X}_{t+1}; Z_{t+1})$, composed of the unique components $Unq(X_t^i; Z_{t+1})$ that are dispersed throughout the solvent (Yurchenko 2024). Thus, the spontaneous growth of a large-scale crystal in a supersaturated solvent gives us an example of self-organized synergy, where causal processes are unambiguously presented by physical interactions.

The question we will be most interested in the next section is this: How are macroscopic emergent phenomena carried out by linear causal chains across scales? Could downward causation be possible due to synergy?

## 4. Synergy and Downward Causation

The PID was not initially developed by Williams and Beer to address causation. It was later suggested that the synergistic component $Syn(X,Y;Z)$ could account for downward causation in stochastic dynamical systems, thereby reconciling the strong and weak forms of emergence (Varley and Hoel 2022). The proof utilizes time-delayed mutual information, interpreted in terms of predictive power, and places it in the context of Integrated Information Theory, which baggage is implicitly based on an assumption that conscious experience is identical to the maxima of integrated information $\Phi$, and can have free will to causally affect the brain, *intrinsically* overcoming its own neural correlates (Tononi et al. 2022).

 doi:10.20944/preprints202509.0678.v1



By replacing the PID-framework with the $\Phi$ID-framework, Rosas et al. (2020) transform the predictive power of a supervenient feature (macroscopic variable) $V_t$ into the causal power of $V_t$ over the underlying dynamical system $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$, presented as above. Thus, the temporal evolution of the system is described by a linear one-body causal chain at a macroscale, where $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t+1}$ serve as the source and target states, respectively. The system is said to have causally emergent feature $V_t$ if and only if $Syn(\boldsymbol{X}_t; \boldsymbol{X}_{t+1}) > 0$. Now, if $V_t$ is associated with particular conscious states, emerging over time from the $\Phi$-structure, while $\boldsymbol{X}_t$ represents the corresponding states of the neural network, this leads to mental (downward) causation defined formally by the following condition (Rosas et al. 2020):

$$Unq(V_t; \boldsymbol{X}_{t+1} | \boldsymbol{X}_t) > 0. \quad (8)$$

In other words, downward causation occurs when an emergent feature $V_t$ has both unique predictive power and irreducible causal power over specific parts of the underlying system. In addition, causal decoupling is proposed when $V_t$ has also predictive and causal properties not only over any specific part but also over the system as a whole (Mediano et al. 2022):

$$Unq(V_t; V_{t+1} | \boldsymbol{X}_t, \boldsymbol{X}_{t+1}) > 0. \quad (9)$$

In fact, what Equation 8 and 9 have shown is that downward causation could be possible if the predictive power of information-based measures about a system, derived from observations, not only reflects causal processes that unfold in spacetime by physical principles, but also, if the system itself is information-processing, becomes equivalent to the causal power of the system itself as expressed in terms of the $\Phi$-ontology. This assumption can be seen as a specific part of a more general hypothesis, called "the hard problem of life" by Walker and Davies (2017), which suggest that a full resolution of the problem of how information, intrinsically processed by living systems, can causally affect matter, will not ultimately be reducible to known physical principles.

Now, consider the dynamical system $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$ in terms of the center of mass, which is the mean location of a body's mass distribution in space. In the case of a system consisting of many bodies, the center of mass is calculated as the average of their masses factored by the distances from a reference point. The center of mass can then be associated with a supervenient variable $V_t$ in Eqs. 9 and 10. Indeed, Rosas et al. (2020) have shown that the center of mass of flocking birds in a 2D computational model predicts its own dynamics via mutual information $I(V_t; V_{t+1})$ better than it can be explained from the behavior of individual birds, i.e., via $\sum_i I(X_t^i; V_{t+1})$. The authors propose this result as an illustration of their theory of causal emergence. The center of mass emerges from the system's dynamics as a gravitational pole that does not physically exist. This point-like center, though computationally powerful, may occupy empty space, having, by definition, no causal power since no observable events might occur there.

Analogously, in thermodynamics, temperature, as the average kinetic energy of particles in a system, is a coarse-grained variable that represents the behavior of all the particles. This supervenient variable allows to predict the system's future state better than it could be made by measuring the speed of individual particles. The predictive value of macroscopic observations is undeniable: the second law of thermodynamics could not even be inferred from observations of microscopic reversible processes in statistical mechanics. However, it does not endow temperature with causal power (even though temperature is conventionally involved in the mechanical work done by a heated system).

Similarly, knowing someone's conscious (supervenient) state at the present moment allows to more accurately predict their future state than knowing their brain's neurodynamics. For example, if someone (say Alice) is asked to choose between an apple and an orange, it can be predicted that Alice's next state $V_{t+1}$ will include either "apple" or "orange" with an equal probability. With additional knowledge about Alice's desires and beliefs, a psychologist might make more precise predictions about her choice, whereas a neuroscientist would hardly reach this level of predicting the future state $\boldsymbol{X}_{t+1}$ of Alice's brain from the previous state $\boldsymbol{X}_t$, both determined by a configuration of activated neurons (leaving aside the problem of decoding brain states into mental states). Note also that "desires and beliefs" are concepts of folk psychology, which can suggest a plausible reason

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 September 2025 doi:10.20944/preprints202509.0678.v1



(explanation) for Alice's choice; however, these concepts could not even be formulated in terms of physical causation, defined on a dynamical system $\boldsymbol{X_t} = (X_t^1, \ldots, X_t^n)$ in a state-space $\mathcal{X}$.

Yet, the choice made by Alice is typically associated with her free will. One could then compute something like a synergistic core in Alice's brain, and identify this statistically well-informed and powerful entity with her Self or conscious "I" capable of exerting mental downward causation on the executive motor modules in Alice's brain (Figure 2).

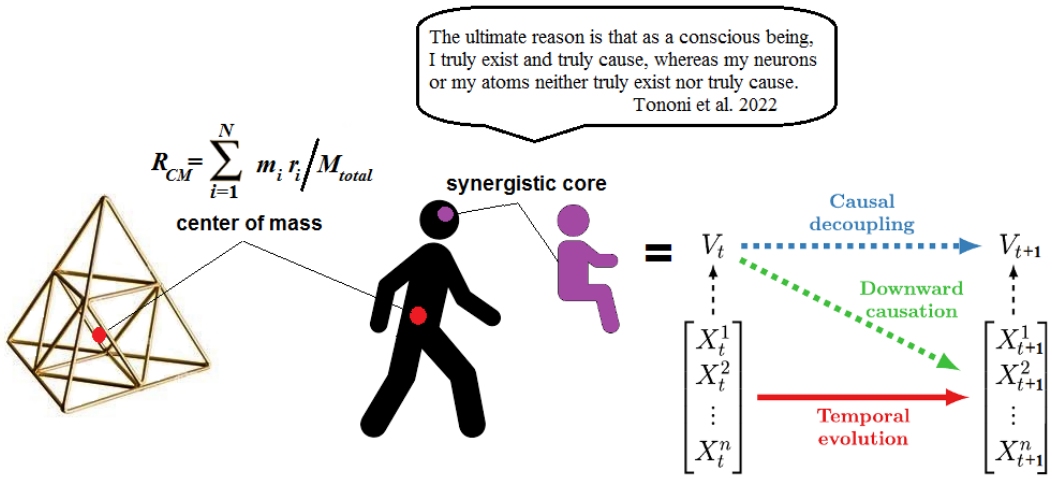


**Figure 2.** Although both the center of mass and the Self can be physically abstract, these concepts provide a useful coarse-grained approximation when applied to systems of interest at larger scales. The former has many applications in mechanics, engineering, and astronomy, while the latter is commonly used in population dynamics, game theory, social sciences, and economics where conscious individuals are modeled as self-interested "black-boxes" interacting freely with each other according to a set of rules.

The Causal Equivalence Principle, discussed in the next section, will disprove this possibility. The aim is to show that consciousness has no more causal power in the brain than temperature does in an ordinary physical system.

## 5. Causal Equivalence Principle

Downward causation requires the strong form of emergence, which is incompatible with reductionism. Reductionism argues that causation is valid only at the smallest scale of physical analysis. This requires a preferred scale for linear causal chains, whereas a common practice in modern science tells us that every science analyzes causal processes at a scale that is most appropriate for the systems of its interest. It would be very difficult or even practically impossible to explain cognitive brain functions at the atomic scale or to develop a sociological theory from a perspective of neural interactions. Reductionists see it as a matter of tradeoff between reasonable simplicity and the scope of detail, especially when fine-grained models may incur prohibitive computational costs, compared to low-dimensional coarse-grained models. This view dominates in physics despite the fact that fundamental physical laws (Newton laws or conservation laws) are scale-independent. Historically, Newton formulated his laws of motion without any knowledge about atoms. Likewise, a neuroscientist can study neural processes at different scales by means of appropriate models such as the Hodgkin-Huxley model, Wilson-Cowan neural mass equations, or Kuramoto coupled oscillators model, with no reference to the atomic scale.

The concept of causation is derived by us from the observation of events. Although each particular event occurs at a corresponding scale, the concept of an event is scale-independent, and no preferred scale can be assigned to it. As indicated above, we refer to completely different things as "events" whether it be a sunrise on Earth, a car accident on the road, a firing neuron in the brain, or

 doi:10.20944/preprints202509.0678.v1



the result of a quantum measurement in a physical lab. Since these systems can be studied at various spatial scales from atoms to whole organisms and large-scale environments, the events they produce occur at corresponding scales. However, it makes no sense to say that the same event will manifest itself at multiple scales, since a microevent cannot in principle be observed at the macroscale, and a macroevent cannot appear at the microscale. On the other hand, changing the scale of observation does not change physical reality, which exists independently of observations.

The scale-independence of causation can be formulated by analogy with the equivalence of all inertial frames of reference in relativity theory, which postulates that the laws of nature are invariant in all inertial frames of reference. Because of this equivalence, observations in one inertial frame can be converted into observations in another frame by the Lorentz transformation with respect to the speed of light. Likewise, one can assert that the dynamics of a system, governed by the laws of nature, cannot depend on the scale of observation.

**The causal equivalence principle (CEP).** *Coarse-grained and fine-grained variables must yield the same dynamics and/or make consistent predictions on the temporal evolution of a system of interest, except for the scope of detail.*

The CEP is not a rule extracted ad hoc from observations, but follows from the inherent properties of spacetime. Its detailed derivation from the causal set $\mathcal{L} = (M, \prec)$ in Minkowski space $(M, g)$ can be found in (Yurchenko 2023a). The proof starts with the microscale and demonstrates how the 'amount' of all linear causal chains available there can be conserved by compressing them in space and time. In relativistic spacetime, spatial compression is consistently provided by mapping all simultaneous and, hence, mutually independent microevents onto their temporal slices, each defined as an equivalence class on a spacelike surface. Accordingly, temporal compression occurs along all timelike worldlines, where each linear causal chain of microevents is condensed into a single pair of microevents, typically based on the temporal resolution of observation provided there. As a result, spatial temporal compression both transform all microevents possible in spacetime into macroevents regardless of their location.

Although the CEP was derived in (Yurchenko 2023a) from the idea of spacetime compression, no event might be observed in empty space. There should be physical systems to produce events as their instantaneous states, which are causally connected in spacetime. Therefore, the "compression" should apply not to spacetime but, rather, to the universe as the largest dynamical system occupying spacetime. In this case, the CEP represents a metaphysical realism: *The existence of the universe is observer-independent*. But how does the universe exist? Is it an atomic (quantum) universe as reductionism claims? We know that organisms consist of atoms, but there are no living entities at the atomic scale. Life is a large-scale emergent phenomenon. In this sense, from a reductionist perspective, the existence of living systems may be viewed as illusory in the atomic (quantum) universe. If so, how do living entities such as bacteria and humans exist? More broadly, how do multiscale dynamical systems like organisms, ecosystems, and planets exist? To answer these questions, we must extend the above postulate to the statement: *The universe, with all its components, exists at all spatial scales simultaneously, regardless of whether or not they are accessible for observation*.

In this context, the CEP can then be generalized as the law of conservation of causation, expressed in the formalism of Liouville's theorem (Yurchenko, 2025). This theorem states that the density of representative points of a dynamical system in phase space does not change with time. Its consequence is that the entropy of the system, defined as the logarithm of the volume in phase space, $H = \log V_\Gamma$, remains constant for a perfect observer capable of distinguishing all causal chains within a system. However, for causal analysis, it is more appropriate to consider the theorem in terms of the continuity equation in fluid dynamics. This equation represents the idea that matter is conserved as it flows in spacetime:

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \mathbf{u}) = 0. \qquad (10)$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 September 2025 doi:10.20944/preprints202509.0678.v1



Here $\nabla$ is the divergence operator, $\rho$ is the flow density, and $\mathbf{u}(\boldsymbol{x},t)$ is the flow velocity in a vector field. For an incompressible fluid, the density $\rho = const$, so that the divergence of the flow velocity is zero everywhere, $\nabla \cdot \mathbf{u} = 0$. Informally, the equation implies that the control volume of a flow remains constant over time (Figure 3a).

Now, let us consider a multiscale dynamical system $\boldsymbol{X_t} = (X_t^1, \dots, X_t^n)$ of $n$ variables. Its macrostate $\boldsymbol{X}(t_0)$ can be observed as a macroevent $E$ which, by definition, is the "sum" of all simultaneous and causally independent microevents $e_i$ at a moment $t_0$, each associated with a corresponding state of the system component $X_t^i$. Suppose there is a linear causal chain of such microevents $e_{i1} \to e_{i2} \to e_{i3} \to \cdots \to e_{in}$ per unite time $\Delta t = t - t_0$. The causal order will still be preserved in a chain of macroevents $E_1 \to E_2$ by reducing the temporal resolution of observations to the lag $1 = \Delta t$. Therefore, the transition across spatial scales does not change the "quantity" of causation within a multiscale dynamical system for a perfect observer (Figure 3b). All scales are causally closed.

Another explanation comes from the principle of locality in physics, which states that an object can be causally affected only by its immediate surroundings and not by distant objects, also known in relativity theory as the dictum "No instantaneous action at a distance," which implies that an action between events is limited by the speed of light. When applied to scale analysis, the principle of locality entails another fundamental property of causation. It states that a microevent $e$ (representing the state of an object at time $t$) at a given scale can be influenced simultaneously by two or more neighboring events (objects), but not by numerous distant events at that same scale. If that were the case, the combined (non-local) effect of these distant events could be regarded as a macroevent $E$ that exerts downward causation on the microevent $e$. Thus, the principle of locality forbids the whole from influencing its individual parts. Causally, the whole cannot be greater than the "sum" of its parts.

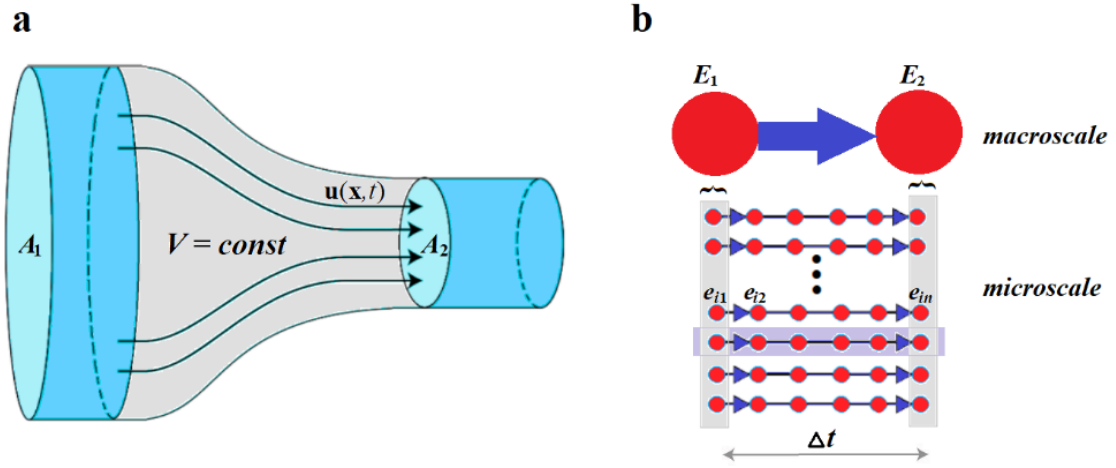


**Figure 3.** (a) The continuity equation states that the control volume $V$ (colored in blue) of an incompressible quantity moving through a pipe remains constant over time. Therefore, as the flow area $A$ reduces, the velocity $\mathbf{u}(\mathbf{x},t)$ increases. It follows immediately that the volume does not depend on how it is measured, i.e., $V = const$ regardless of the units of observation. (b) Here, the control volume is schematically represented by $n$ linear causal chains passing through spacetime occupied by the dynamical system $\boldsymbol{X_t} = (X_t^1, \dots, X_t^n)$. Accordingly, all simultaneous microevents (a vertical bar) indicate the flow density, i.e., the number of causally independent microevents $e$ per unit volume, whereas each row of causally connected events (a horizontal row) at the microscale corresponds to the flow velocity vector, i.e., the average number of microevents $e$ per unit time in the linear one-body causal chains. For a perfect observer, the quantity of causation within a control volume of spacetime remains constant across scales. No cause can appear *ex nihilo* at a particular scale in order to intervene in linear causal chains with their own past.

 doi:10.20944/preprints202509.0678.v1



**The law of conservation of causation.** *The flow of causation in the universe is conserved across scales.*

In the context of Noether's theorem, for every continuous symmetry in the laws of nature, there exists a corresponding conservation law. Accordingly, conservation of causation can be inferred from the invariance of the laws of nature under scale transitions. In effect, this law states that the choice of units of observation in spacetime does not affect the flow of causation. No linear causal chain at a fixed scale can intervene in linear causal chains at other scales. The law rules out both upward and downward causation, which can only appear as artifacts of imperfect observation when different scales of causal analysis are mixed. On the other hand, according to the law, the Markov property remains invariant across scales due to the linearity of causal chains.

## 6. Corollaries of the CEP

The CEP has corollaries that are directly relevant to the relationship between causation, information, and synergy in biological systems discussed in the previous sections.

**Corollary 1.** *There is no causally preferred spatial scale.*

Thus, the CEP is not a reductionist principle. In physicalist terms, reductionism is based on two premises: (i) micro-causal closure and (ii) macro-causal exclusion (Kim 2006). In contrast, the law of conservation of causation states that not only the microscale is causally closed, but every scale is causally closed.

Formally, the CEP is similar to the Principle of Biological Relativity of Noble et al. (2019) which states that there is, *a priori*, no preferred level of causation across the multiple scales of networks that define the organism. However, there is a principled distinction between them. Biological relativity has no relevance to relativity theory, and typically conflates the notion of reason (as a logical cause-like explanation) with the rigorous concept of cause (as a physical event in spacetime, linked to an instantaneous state of a system of interest). As a result, this admits cross-scale causation. Upward causation is defined by the mechanics that describe how lower elements in a system interact and produce changes at higher levels. Downward causation is represented by the set of constraints imposed by environmental (large-scale) conditions on the system's dynamics at lower levels.

In contrast, the CEP forbids both kinds of cross-scale causation. Corollary 1 is compatible with Rolls' approach to causation, which argues that (linear) causal chains operate within scales but not across scales. He regards downward causation as a philosophical "confabulation" aimed at disproving reductionism (Rolls 2021).

**Corollary 2.** *Mixing two (or more) causally closed scales leads to the double causation fallacy or overdetermination.*

Note that "overdetermination" here must not be confused with a case when two or more simultaneous events cause another event, all occurring at the same scale. Such collisions (and bifurcations) between different linear causal chains can be ubiquitous in the causal set $\mathcal{L}$. Corollary 2 permits linear causal chains to intersect at any one scale but not across scales. The double causation fallacy arises when both macroscopic and microscopic variables are supposed to causally affect the same event (Figure 4a) despite the fact that microevents cannot in principle be observable at the macroscale, and vice versa.

In the classic example of overdetermination, as suggested by Kim (2006), there are two emergent mental states M and M* that supervene on physical states Q and Q* of a system S respectively. Now, if we agree that M causes M*, then we must also agree that Q is causal for M*. If both M and Q explain M*, then the explanation is overdetermined. This example, however, is more concerned with the mind-body problem than with downward causation. As stated, assuming downward causation alone is not sufficient to solve the mind-body problem. To account for mental causation, neo-Cartesian dualism is necessary to conflate causation with information.

 doi:10.20944/preprints202509.0678.v1



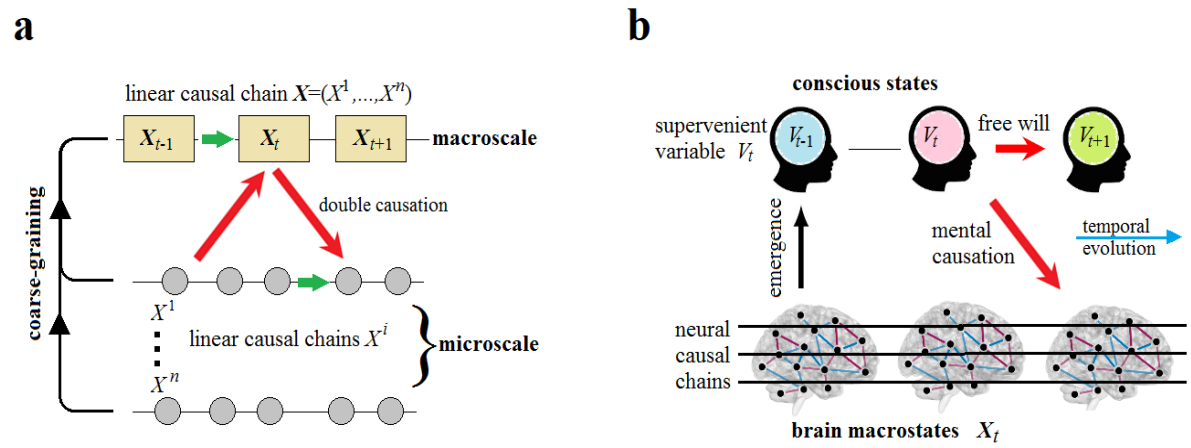


**Figure 4.** (a) Coarse-graining allows to reduce data from a high-dimensional space to a low-dimensional space by aggregating many microscopic variables into a single macroscopic variable. This is modeled by Markov chains, which are implicitly derived from observations of events in linear causal chains. Thus, coarse-graining 'compresses' many causal chains $X_t^1$ at the microscale into a single causal chain $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$ at the macroscale. Causal chains evolve at a corresponding scale through their own cause-effect relations (shown via green arrows). The CEP ensures that causal chains only intersect at the same scale and not across scales. The double causation fallacy arises if the temporal evolution of a chain at one scale is assumed to be affected by a chain from another scale (indicated by red arrows). (b) Here, the stream of conscious states, represented by a supervenient variable $V_t$, emerges from the brain temporal evolution. Neural causal chains pervade the brain at the microscale, whereas brain dynamics are represented by macrostates $\boldsymbol{X}(t)$. Now, if information is conflated with causation, downward causation turns into mental causation, affecting neural causal chains. Accordingly, at the macroscale, mental causation takes the form of free will, influencing one's conscious states (which are ultimately responsible for one's behavior in the environment).

To specify the problem, we translate the above example into the formalism of multiscale dynamical systems. Let large-scale supervenient variables $V_t$ and $V_{t+1}$ represent mental states M and M* of a multiscale dynamical system $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$ such as the brain. The temporal evolution of each its part $X_t^i$ can be represented by a linear causal chain at a corresponding scale less than the scale of the system. In dynamics over time, the state $X_{t+1}^i$ of each part is determined by its previous state $X_t^i$ (though other parts can intervene in the chain). Double causation arises when $X_{t+1}^i$ is also affected across scales by $V_t$. Thus, mental causation would be possible if the conditions of Equation (8) were satisfied (Figure 4b). But this is not the case for the CEP.

Ultimately, the CEP rejects the strong version of emergence (including downward and mental causation) but adopts its weak version. The CEP may maintain a conventional form of free will on the condition that brain dynamics could not be completely predetermined from the past. Let $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t+1}$ be two brain states of a corresponding one-body linear causal chain at the macroscale. We say that $\boldsymbol{X}_{t+1}$ is caused by $\boldsymbol{X}_t$. Now let $V_t$ symbolize this linear chain that can be described as a discrete stream of conscious states emerging in critical points of Langevin dynamics (Yurchenko 2023b). According to the CEP, all scales are causally closed so that mental (downward) causation from $V_t$ to $\boldsymbol{X}_{t+1}$ is precluded. Instead, conscious states passively emerge from brain macrostates. This can be formally provided by mapping the brain states $\boldsymbol{X}_t$ to the corresponding conscious states $V_t$ in the stream. Thus, the causal relationship between $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t+1}$ is spontaneously transformed via the mapping into the mental relationship between $V_t$ and $V_{t+1}$ as if the former caused the latter (Figure 5).

 doi:10.20944/preprints202509.0678.v1



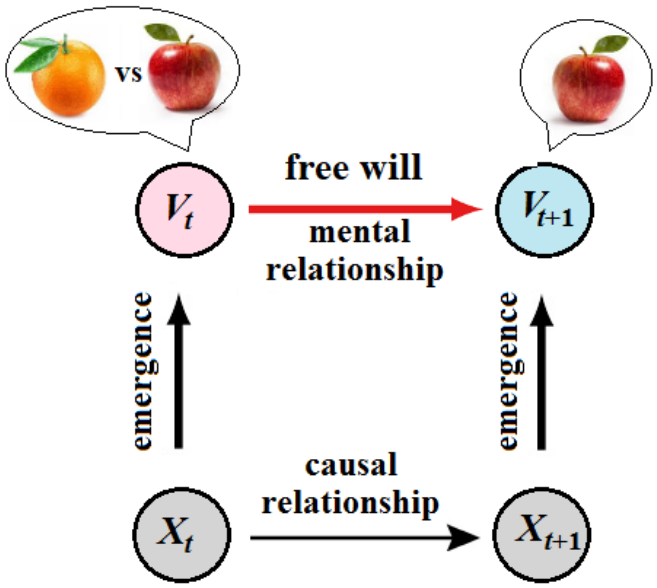


**Figure 5.** The CEP allows us to account for the conventional form of free will by converting a physical cause-effect relationship between two nearest states of the brain into a mental relationship between two corresponding psychological states. Although this formal mapping between physical brain states and subjective conscious experience does not solve the mystery of their relation, this can explain the illusion of free will, while preserving the ability of the brain or, more generally, conscious organisms to act freely.

**Corollary 3.** *Endogenous coarse-graining is causally legitimate.*

Here, coarse-graining is not related to dimensionality reduction in statistical analysis such as principal component analysis or data compression. Instead, it concerns actual events as they occur according to natural laws, forming a lattice of causal chains in spacetime. Coarse-graining depends on the spatial and temporal resolution of observations as if spacetime itself was compressed, making only macroevents observable. Examples of coarse-graining include temperature in thermodynamics, the center of mass in mechanics, condensed nodes in network analysis, and phase-space models that reduce population size to a single variable in population dynamics. These examples illustrate weak emergence, where "the map is better than the territory" (Hoel 2017), rather than strong emergence, where "the macro beats the micro" (Hoel et al. 2013).

Flack (2017) suggests to distinguish this kind of coarse-graining, imposed by scientists on a system of interest to find compact descriptions of its behavior for making good predictions, and *endogenous* coarse-graining, imposed by nature itself upon matter. Endogenous coarse-graining is what allows us to distinguish between a physical body and its environment. A physical body is by definition a system of components that are more causally connected to each other than to the components of the environment. In the case of an atom, its electrons are more causally connected to each other and to the atom's nucleus than to the electrons and nuclei of surrounding atoms. Similarly, a molecule consists of atoms that are chemically coupled with each other either directly or through causal linear chains to a degree that exceeds their coupling with the atoms of other molecules. At a larger scale, the surface of a biological cell is a boundary between the causal strength of internal (atomic and molecular) interactions within the cell and its interactions with the environment.

Endogenous coarse-graining is evident in the multiscale organization of matter from atoms to planets, and, especially, in biological systems, organized across scales from genes to cells to organisms. Although linear causal chains pervade spacetime uniformly, physical bodies and living organisms are more internally connected than externally. This makes them partially autonomous from their environment. In fact, internal causal connectedness (or causal closure) is fundamental to our understanding of physical existence. For example, when we see a stone rolling down a hill, we

 doi:10.20944/preprints202509.0678.v1



perceive the stone as a distinct physical body separate from the parts the hill consists of. Also, through observations we can distinguish a leaf on a tree and a tree in a garden across different spatial scales. Why are we sure that these are real physical entities and not illusions created by the sophisticated computational power of the brain, which transforms the beam of photons hitting one's retina into a series of visual images?

This problem can be traced back to the words of the French philosopher Hippolyte Taine: "Instead of saying that a hallucination is a false exterior percept, one should say that the external percept is a true hallucination" (Corlett et al., 2019). In this context, our observations of hills, stones, gardens, trees, and leaves are true epistemologically because they all exist as endogenously coarse-grained entities. The brain, as a predictive system, could not obtain the information necessary to discern them if they were not internally more causally connected than externally. Endogenous coarse-graining here means that these entities exist ontologically due to their internal causal connectedness, not just as artifacts of our observation. In fact, internal causal connectedness is the primary, if not the only, intrinsic property of physical entities that enables us (our brain) to distinguish them and their parts from one another and from their environments.

Endogenous coarse-graining is closely related to the concept of individuality in biology, which is typically divided into four kinds: metabolic, immunological, evolutionary, and ecological individuality (DiFrisco 2019; Kranke 2024). Since information is physically instantiated in the organizational structure of matter and conveyed through spacetime causally (Figure 1a), the relationship between endogenous coarse-graining, based on internal causal connectedness, and biological individuality can be statistically inferred in terms of time-delayed mutual information or transfer entropy from the definition: "If the information transmitted within a system forward in time is close to maximal, it is evidence for its individuality" (Krakauer et al. 2020). A similar measure, based on the concept of non-trivial information closure (Bertschinger et al. 2008), is proposed in neuroscience to explain the large-scale emergence of consciousness (psychological individuality) from neural activity in the brain (Chang et al. 2020). Thus, endogenous coarse-graining, as an intrinsic property of matter organization across scales, underlies a set of interdependent concepts in biology and complex systems such as emergence, synergy, self-organization, information closure, autonomy, biological individuality, autopoiesis, cognition, and subjective experience.

Although the law of conservation of causation explains why different levels of description of the same system can co-exist and be causally valid, the question we are interested in now is how the scales are related to each other within the spatial span of endogenously coarse-grained dynamical systems.

## 7. Causal Scope, Scale Transition, and Spatial Span

In practice, scientists are naturally constrained to choose an *elementary basis* for the lowest boundary of observation and causal analysis at a scale that is most suitable for the size and dynamics of a system of interest (e.g., the solar system versus a cell). The basis can be chosen explicitly or implicitly, but it will always be embedded in the framework of research. All scales below the basis are ignored (e.g., quantum, atomic, molecular). There will also appear the upper boundary of observation for causal analysis over the spatial (and temporal) span of the system of interest. Again, all scales above the upper boundary are ignored and commonly related to the environment (e.g., populational, ecological, planetary). Within this span, three scales are typically proposed: the microscale for the elementary basis, the macroscale for the system itself, and a mesoscale between them.

At first sight, graph theory provides the best representation of causal chains that occur and are valid only at the same scale of spatial resolution. However, the graph, defined on a set of nodes, does not discriminate between scales: one node is taken to be of the same size as another node. Although the graph can then be coarse-grained by condensing local networks of strongly interconnected nodes into single large-scale nodes (modules), such transformations would change the scales of description but preserve one-scale representation. According to the CEP, the graph-theoretic representations

 doi:10.20944/preprints202509.0678.v1



could indeed be best for dynamical causal modeling, but only on the condition that all the nodes were physically related to the elementary basis of a system of interest. In the case of the mind-brain relation, the elementary basis should be related to the scale of single neurons, while communication channels for linear causal chains between neurons should naturally be provided by structural (anatomical) connectivity via white-matter fibers. Unfortunately, graph-theoretic representations can roughly mix different spatiotemporal scales, as it occurs by detecting statistical correlations of functional connectivity between different regions of the brain with the help of various neuroimaging techniques. Spurious causation can arise there (Reid et al. 2019; Weichwald and Peters 2021; Barack et al. 2022).

Overall, the CEP argues that observers should keep the scales of causal analysis isolated. Theoretically, we should denote the elementary basis $N$ of a system of interest as scale 1. The sets $S$ of $n$ elements, $S \subseteq N$, over the basis, should spontaneously produce the class of equivalence by their cardinality, assigned then to scale $n$. Each new scale $n+1$ would occur by adding a new element to each of the sets. Scales would be additive, in the sense that a set of interdependent components might be replaced by a single component whose scale is equal to the sum of the scales of the individual components (Allen et al. 2017). In practice, however, the distinction between scales cannot be so simple. For example, where should we place the boundary between the microscale and a mesoscale in our observations? More broadly, how should the molecular scale transition spontaneously into the cellular scale, and how should the scale of single neurons consistently turn into the scale of neural functional modules?

This problem is very similar to the so-called "heap paradox" in philosophy, which argues that if a grain of sand is not a heap, and adding a single grain of sand to something that is not a heap does not produce a heap, then a heap is physically impossible despite the fact that it *emerges* in the eye of an observer. Does the heap of sand really exist? The answer must start with the remark that every grain of sand is already a "heap" of atoms more causally connected with each other than with atoms of other grains. So, a heap of sand arises at a larger scale analogously when a number of grains become causally coupled with each other stronger than with the environment. Therefore, intervention on any part of the heap will causally affect other parts rather than distant objects outside the heap, for example, to bring about an avalanche.

What follows from this dilemma is that the boundaries between scales cannot be defined rigorously but only in terms of neighborhoods, by analogy with topological spaces, in which closeness (or limit) is described in terms of open subsets rather than metric distance. In the topological presentation, the number of interacting individuals must be fixed, whereas in the metric presentation, the number can vary with density per unit volume. In a graph $G = (N, E)$, let each unit scale $u_i$, defined as the equivalence class by cardinality $|k|$, be assigned to the neighborhood $O$ such that $1 \leq k \leq O$, where $k$ is the causal scope (similar to the degree of a node or branching factor in graph theory), which is specific to a particular class of systems, and $O$ determines the topological scale parameter, imposed by the principle of locality that is fundamental to causal analysis.

Now, the CEP allows an event to cause $k$ consequent events or, equivalently, to be caused by $k$ previous events at the same scale of observation. However, the principle of locality forbids $k$ to be larger than the neighborhood $O$ of the given scale. The nearest upper scale is defined as $u_{i+1} \stackrel{\text{def}}{=} |u_i k|$. Thus, the progression of scales will grow exponentially, starting with the initial unit scale $u_1 = k$ in the elementary basis of $G = (N, E)$ (Figure 6a):

$$(\forall i) u_i \stackrel{\text{def}}{=} k^i. \qquad (11)$$

It is important to note that Equation (11) must be interpreted correctly. The causal scope does not imply that a system is divided into groups of $k$ elements, each forming a clique isolated from the rest of the system. If that were the case, scale transitions would only increase the size of cliques (Figure 6b). Instead, while the groups are defined numerically by the equivalence class, their neighborhoods $O$ topologically cover the system entirely at every scale by involving intersections between nearest groups with each other, rather than partitioning it into isolated groups of

 doi:10.20944/preprints202509.0678.v1



exponentially growing size. Therefore, scale transitions increase the coherence of the system as a whole without making it a clique of completely interconnected elements.

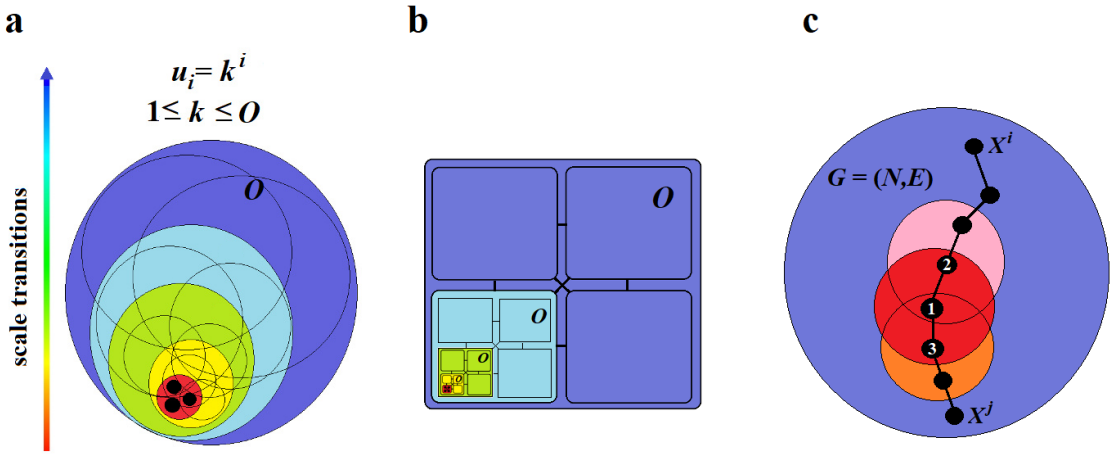


**Figure 6.** (a) Although the concept of "scale" is fundamental in science, it is impossible to define each scale rigorously. Causal analysis can hardly be based on a metric scale length when dealing with multiscale systems. In that case, scales can be defined topologically by neighborhoods $O$ of limited size, each of which determines a characteristic scale parameter. (b) The causal scope of two elements defined numerically by an equivalence class $|k|$, does not coincide topologically so that their neighborhoods $O$ do not form cliques of exponentially growing size across scales. (c) At every scale, these neighborhoods cover the whole system by intersecting with each other. So, nodes (events) 2 and 3 may be in the neighborhood of node 1, but their own neighborhoods do not coincide. Coherence of a dynamical system $\boldsymbol{X}_t = (X_t^1, \dots, X_t^n)$, represented by a graph $G = (N, E)$, means that there is a linear multi-body causal chain of finite length $L$ connecting any two variables $X_t^i$ and $X_t^j$ over time.

Coherence here refers to the so-called small-world property of a graph, also known as the six degrees of separation in social networks. This property is characterized by a high degree of local clustering and a relatively short path length. The latter is the average number of edges in the shortest path between two nodes in the graph $G = (N, E)$, defined as $L = \ln N / \ln k$ (Watts and Strogatz 1998). In other words, the internal causal connectedness of an endogenously coarse-grained system $\boldsymbol{X}_t = (X_t^1, \dots, X_t^n)$ guarantees that although the principle of locality forbids its two elements at a distance to have a direct causal link, they can still correlate over time via a multi-body causal chain of finite length of mutually causally connected neighbors within the system (Figure 6c). We can thus be certain that such a causal chain exists or is intrinsically feasible between any two organelles in a cell, two neurons in a brain, or two organs in an organism.[1]

Now, let us consider the multiscale nonlinear phenomena of coherence, such as the spontaneous growth of a crystal in a supersaturated solvent or an avalanche that can occur in a heap of sand, a flock of starlings, or a neural network in the brain. These phenomena are triggered by a single element such as a seed crystal, a grain of sand, a single bird, or a neuron. One might argue that these dynamics exemplify upward causation, enabling a part to causally affect the whole across scales. Another

[1] In particular, this raises an interesting question about biological individuality in the context of social small-world networks with their six degrees of separation. If each degree of separation can be viewed as a causal link provided via a communication channel between two people, then the chain of acquaintances represents the shortest path length $L \cong 6$ that "causally" connects any two humans on Earth. Does this mean that humanity can be regarded not only as a taxonomic species, but also as a single biological individual, i.e., a genuine superorganism? Or, is this a case of confusing causation with information? Putting it in the context of folk psychology, should we agree that a word spoken by one person to another person has causal power over the latter similar to that between two organelles in a cell or two neurons in the brain?

 doi:10.20944/preprints202509.0678.v1



famous example of upward causation is the butterfly effect, a metaphor in chaos theory, where a microevent, the flap of a butterfly's wings, can ultimately cause a series of macroevents, like a hurricane. But is there upward causation?

In a dynamical system of $N$ elements, each element can only causally affect up to $k$ elements at time $t$, but it cannot simultaneously influence an unlimited number of elements beyond its neighborhood $O$ as dictated by the principle of locality. For example, for a starling flock, the causal scope $k \cong 7$ (Ballerini et al. 2008), whereas for the rat barrel cortex, $k \cong 28$ (London et al. 2010). According to the law of conservation of causation, when the scope of elements exceeds $O$, the scale of observation must be changed (Figure 6a). This marks the boundary between the two closest scales over which the avalanche progresses, making a system more coherent over time. Instead of upward causation, there is a great number of linear multi-body chains with a common cause in the distant past that triggered the avalanche. There are statistical (functional) correlations among all elements of the avalanche due to the common cause, but there is no immediate causal link between them. If there were, we would have a brain where all neurons form a clique completely interconnected by white-matter fibers. Thus, this is not the case.

Remarkably, assuming upward causation in the above scenario would make downward causation possible as well. If one element was capable of affecting an unlimited number of elements simultaneously, then the opposite process should occur spontaneously. Eventually, there would be a moment when a large number of elements simultaneously affect one element, as if the whole could causally impact its parts. This scenario implies that the brain should again be a clique of neurons completely interconnected via synaptic communication channels despite neurobiological evidence indicating that in the brain, each neuron has, on average, several thousand synaptic connections with other neurons, and only a few can be activated simultaneously. This evidence also demonstrates how the principle of locality is instantiated in the anatomical connectivity of the brain.

The spatial span of an endogenously coarse-grained system can now be defined not as the metric volume in space occupied by the system, but as the number of spatial scales causally covered by its dynamics. According to Equation (11), the spatial span $S(G)$ of a network $G = (N, E)$ can be calculated logarithmically by its elementary basis $N$ as $\log_k N$. For example, if we consider the human brain, which contains approximately 86 billion neurons, and assume that the causal scope of a neuron, as detected experimentally by single-cell stimulations (Kwan and Dan 2012), is $17 \leq k \leq 23$, then the spatial span of the human brain can encompass 7 or 8 scales, beginning with individual neurons in the elementary basis. For a flock of starlings, assuming it consists of about 10 thousand birds, with $6 \leq k \leq 7$ (Ballerini et al. 2008), its spatial span would range between 4 and 5 scales (Figure 7a). Of course, in both cases, the number of scales would increase significantly by shifting the elementary basis to the atomic scale. Note also that the value of $k$ may vary across scales even within the same system. At the molecular scale, the causal scope, influenced by the topological scale parameter $O$, is likely to differ from that at the neural scale.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 September 2025 doi:10.20944/preprints202509.0678.v1



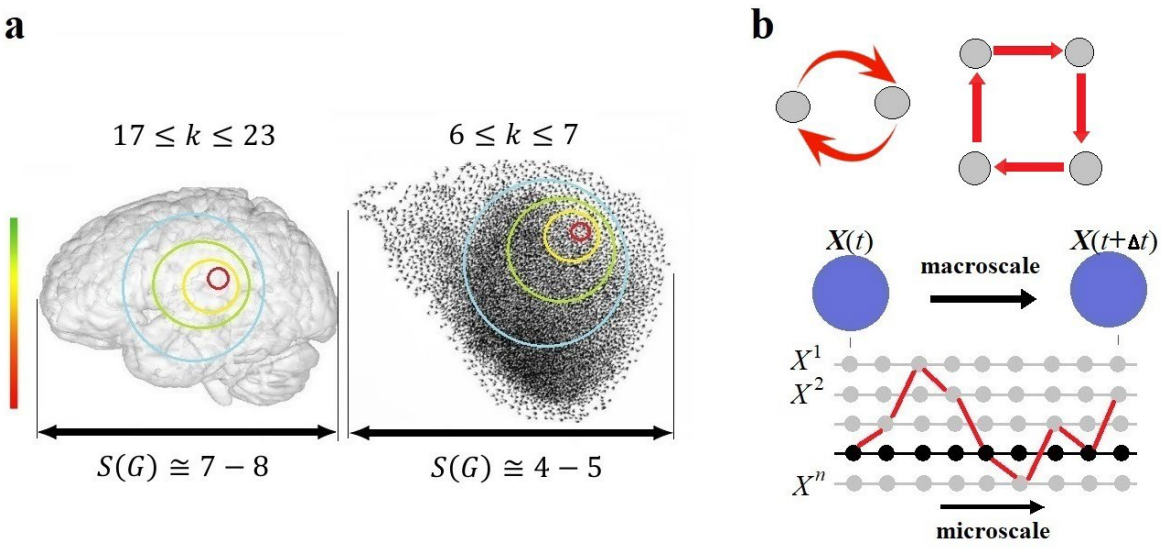


**Figure 7.** (a) The spatial span $S(G)$ of the human brain and the starling flock, both represented by a network $G = (N, E)$. (b) Top: Circular causation is typically depicted in Boolean networks by closed loops connecting two or more nodes. The temporal dynamics are missed in such representations, using classical timeless logic. Bottom: In contrast, time is implicitly embedded in linear causal chains of dynamical systems. Circular causation occurs when a multi-body chain (a red polygonal chain) intersects a fixed one-body chain (a black straight line) many times. If these intersections repeat regularly, causation becomes cyclic. Circular causation can arise independently at different scales.

Another important feature of scale transitions is that they can essentially change the causal structure of a system at each scale. For a dynamical system $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$, the causal scope means that a variable $X_t^i$, associated with a microevent at time $t$, can simultaneously affect $k$ other variables, generating multi-body causal chains. Circular causation arises when one (or more) of the affected variables causally impact $X_t^i$ at time $t + \Delta t$ (Figure 7b). However, circular causation at a smaller scale may not be observable at a larger scale, and vice versa. Macroscopic systems can, thus, exhibit nontrivial complex behavior that could not be inferred from their microscopic components. The CEP can explain how complex nonlinear phenomena can emerge on multiscale networks from linear causal chains solely due to scale transitions, e.g., when a system constrains itself to move through a cyclic attractor in phase space. In this context, the CEP underlies the renormalization group formalism in condensed matter physics, in terms of critical phenomena on Ising models that are typically characterized by spontaneous avalanches across scales (di Santo et al. 2018; Lombardi et al. 2021).

As stated, the CEP is not a reductionist principle. From the reductionist perspective, we should ultimately agree that all endogenously coarse-grained systems, including ourselves, have no biological individuality because only atoms (or quanta) genuinely exist, and have causal power. In contrast, the CEP asserts that all scales are ontologically valid and causally closed. According to the law of conservation of causation, scale transitions preserve the quantity of causation invariant in dynamics when a system passes from one state to another regardless of the scale of observation. However, observations of the system's causal structure at the macroscale do not allow one to uncover its causal structure at the microscale, and vice versa, so reduction is precluded (Figure 8).

 doi:10.20944/preprints202509.0678.v1



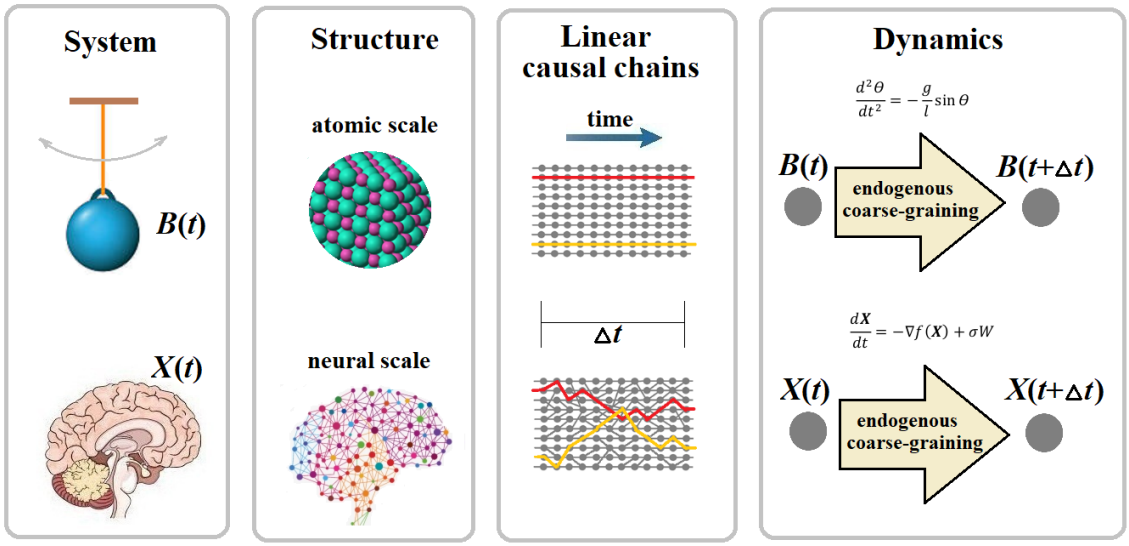


**Figure 8.** The schematic illustrates a comparison between the dynamics of a classical pendulum and the brain, both presented as internally causally connected systems. The pendulum is depicted as a ball $B(t)$, that swings back and forth around the equilibrium position. Given its structure at the atomic scale, all atoms in its crystal lattice move regularly, producing the same one-body linear causal chains in dynamics without intersections between two arbitrary chains (shown as red and yellow lines). In contrast, the brain, depicted by the macroscopic variable $\boldsymbol{X}(t)$, has a highly complex structure, composed of networks at the neural scale. Neurons interact in dynamics, producing multi-body branching and circular linear causal chains that are entangled with each other. Although in macroscopic dynamics the difference between $B(t)$ and $\boldsymbol{X}(t)$ is obscured by endogenous coarse-graining that yields a simple one-body causal chain, going from one state of a system at time $t$ to its next state at time $t + \Delta t$, these consolidated linear chains are very distinct at their characteristic microscale. In fact, every linear one-body causal chain associated with the successive states of a dynamical system $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$ can contain various multi-body causal chains resulting from interactions between the system's components $X_t^i$. Thus, the complex behavior of endogenously coarse-grained systems can increase exclusively due to their multiscale organization. The significance of entanglement and circularity via synaptic fibers becomes especially notable if all neurons in the brain were completely disconnected. In that case, brain dynamics would resemble a causal picture of the pendulum ball.

On the other hand, the CEP can explain the problem of indeterminism for living organisms in the context of their biological freedom to respond to external stimuli without resorting to mental downward causation (Figure 5). First, causation is not synonymous with determinism which heavily relies on the idea of predictability as seen from the perspective of a perfect observer like an omniscient Laplacian demon. Every event is said to be completely predetermined from the past so that randomness (stochasticity) only appears to an observer lacking perfect knowledge. In contrast, causation requires only irreflexivity and transitivity by Equations (1.1) and (1.2). No event can be a cause of itself as every event, associated with the state of a system, is necessarily preceded by other events, including its previous state. Even if the dynamics of organisms are entirely deterministic at the atomic scale, their macroscopic behavior can still have its own causal structure that may not be derived from the causal structure at the microscale. Meanwhile, the law of conservation of causation tells us that there is nothing new added to the "quantity" of causation at the scale of organisms compared to the "quantity" of causation they possess at the atomic scale within their spatial span (Figure 7a). Since their behavioral response to external stimuli is not reducible to atomic interactions, it cannot in principle be proven (or disproven) that the response has been predetermined from the past. According to the CEP, all causal structures within the system's spatial span are consistent so that our free will can be preserved without involving mental downward causation or quantum indeterminism.

 doi:10.20944/preprints202509.0678.v1



So, we may be puzzled when observing systems as different as a simple pendulum and the human brain. The dynamics of both are summed up at the macroscale to a consolidated one-body causal chain despite a huge difference in the complexity of their causal structures at lower scales. (Figure 8). Thus, multiscale hierarchical organization provides not only the biological individuality of organisms based on their internal causal connectedness but also their biological freedom to adapt to the environment. Their ability to adapt is selectively encoded over evolutionary time-scales in the entangled and circular multi-body causal chains pervading their spatial span at all scales (Figure 7b), while atoms, the organisms consist of, do not adapt but follow their deterministic ways.

The following sections will present a mathematical demonstration of how the CEP can be applied to multiscale, hierarchically organized networks.

## 8. Hierarchical organization of Complex Systems

Hierarchy is a universal feature of complex systems in social sciences, biology, and neuroscience (Mihm et al. 2010; Kaiser et al. 2010; Deco and Kringelbach 2017; Hilgetag and Goulas 2020). In causal analysis, however, this term must be taken with caution since there are two very distinctive types of hierarchy. It follows from the fact that "hierarchy" can be conceptualized in two relevant but mathematically different ways.

### *8.1. Flat Hierarchy*

In literature, hierarchy is typically defined as a set of elements (nodes) arranged into ranks or layers. In its most general mathematical formulation, a hierarchy is an acyclic directed graph $G = (N, E)$, also represented as an upper semilattice $\mathcal{H} = (N, \leq)$, where $\leq$ symbolizes subordination in the usual mathematical sense of order. A canonical example is a power hierarchy, which consists of a central authority that is transferred down across subordinated ranks to exert the chains of command and control. The order arises spontaneously within the hierarchy across ranks. In general, given a set $N$, the number of ranks (the height of the hierarchy) is determined by the degree of branching, i.e., by the number of subordinates each node has on average. The $\mathcal{H}$ can be decomposed by linear chains, consisting of subordinated nodes over all ranks in the hierarchy. But such a representation produces great confusion that is responsible for assuming downward causation in complex hierarchical systems.

First, this definition only characterizes the simplest form of a hierarchy where all layers are presented at the same scale. A much more complex example is a nested hierarchy which is composed of subsystems that, in turn, have their own subsystems, and so on. The nested hierarchy is a multiscale modular structure that is *nearly decomposable* (Simon 1969). Simon argued that near-decomposability (modularity) is a pervasive feature of natural complex systems because it provides the emergence of complexity from simple systems through stable intermediate functional modules that allow the system to adapt one module without risking the loss of function in other modules (Meunier et al. 2009). Importantly, unlike a one-scale (or flat) power hierarchy where nodes (elements) are distributed across layers with one node at the top, which is superior to all others, and those at the bottom, which are inferior to all others, in a multiscale hierarchy all nodes are uniformly placed in the lowest layer (scale) as the *elementary basis* above which modules unfold.

### *8.2. Multiscale Modular Hierarchy*

The multiscale hierarchy is a power-set $\mathcal{P}(N) = \{A | A \subseteq N\}$ that can be mapped onto an upper semilattice $\mathcal{H} = (N, \leq)$ by condensing all subsets (including one-element subsets) of $N$ as nodes with no interior content at a given scale: $A \subseteq B \rightarrow \{A\} \leq \{B\}$. Mathematically, the multiscale hierarchy can be defined as an ideal $\Delta$, a structure on the elementary basis $N$, all subsets of which satisfy the following conditions:

$$\begin{cases} A, B \in \Delta \Rightarrow A \cup B \in \Delta; \\ A \in \Delta \ B \subseteq A \ \Rightarrow \ B \in \Delta. \end{cases} \quad (12)$$

 doi:10.20944/preprints202509.0678.v1



In the biological context, the closure of the hierarchy $\Delta$ from above by the conditions (12) establishes the boundary between an endogenously coarse-grained system and its environment as a necessary prerequisite for its self-organization and autonomy. In the neuroscientific context, the closure is also a prerequisite for the existence of the global workspace (Dehaene and Naccache 2001) and non-trivial information closure (Chang et al. 2020), both associated with the emergence of consciousness in the brain. It is important to note that closure is a universal property of $\Delta$ independent of the size of its elementary basis, such as the number of neurons in a neural hierarchy: both the human brain and the mouse brain are hierarchies closed from above. Thus, this property must be present across species (whereas cultures of neurons or slices of cortex from in vitro experiments lack it).

Another universal property of $\Delta$ is its self-similar (fractal) architecture or scale-invariance, viewed as one of the fundamental features of hierarchy. Mathematically, any closed subset $A \subset N$ in the elementary basis of a hierarchy $\Delta$ can spontaneously generate its own sub-ideal, $\Delta_A \subset \Delta$. This makes hierarchy a universal scale-independent phenomenon of nature that can spontaneously emerge over any set of physical units that are causally (and informationally) connected. Thus, both closure from above and self-similarity of hierarchy are natural prerequisites for concepts such as biological individuality (Krakauer et al. 2020), which can evolve at any level of organization and be nested across scales.

In network science, a variety of different measures are suggested to detect connected populations in networks (Rubinov and Sporns 2010; Lynn and Bassett 2019). Unfortunately, the words “level”, “layer” and “scale” are often used interchangeably in the literature. This terminology confuses causal analysis of complex systems. Here and below, these terms will be strictly separated. While “scale” will obviously mean spatial (or temporal) scales arranged logarithmically according to Equation (11), the term “layer” will exclusively refer to the structural organization of a system, studied at the same scale of observation. So, the power hierarchy, the cortical hierarchy of pyramidal cells, or an artificial input-output neural network will all be called *multilayer* as they consist of many functionally subordinated layers, located, however, at the same spatial scale (Figure 9a). In contrast, a hierarchy starting from gene and protein networks to neuronal modules to the whole brain network will be classified as *multiscale* (Figure 9b).

Finally, the term “level” is often used in philosophy to articulate scale-dependent concepts such as mechanistic constitution between parts and wholes, levels of selection, levels of biological individuality, upper-level autonomy, and downward causation (Brooks et al. 2021). On the other hand, in network science, this term implies different conceptual representations of the same system, without reference to scale or layer. A typical example is a multilevel hierarchy formalized as an edge-labeled multigraph (Kivelä et al. 2014; Gysi and Nowick 2020) where all the levels represent the same elementary basis while elements (nodes) in each level are connected by various features of interest (e.g., shape, color, age, gender, family ties, affiliation, skills, and other categories). Another familiar example of different types of interactions is the relation between structural (causal) and functional (statistical) connectivity, which can be represented by two interdependent levels (Signorelli et al. 2022). In other words, the term “level” should be reserved for categorical analysis of qualitative data, such as classifications based on shared characteristics, but avoided in multiscale analysis. Although multilevel decomposition can provide powerful mathematical tools for investigating relationships between elements of networks organized into groups by different kinds (e.g., species in biology), these levels cannot be ordered by scales or even by subordinated layers. Strictly speaking, multilevel networks should not be called hierarchical at all.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 September 2025 doi:10.20944/preprints202509.0678.v1



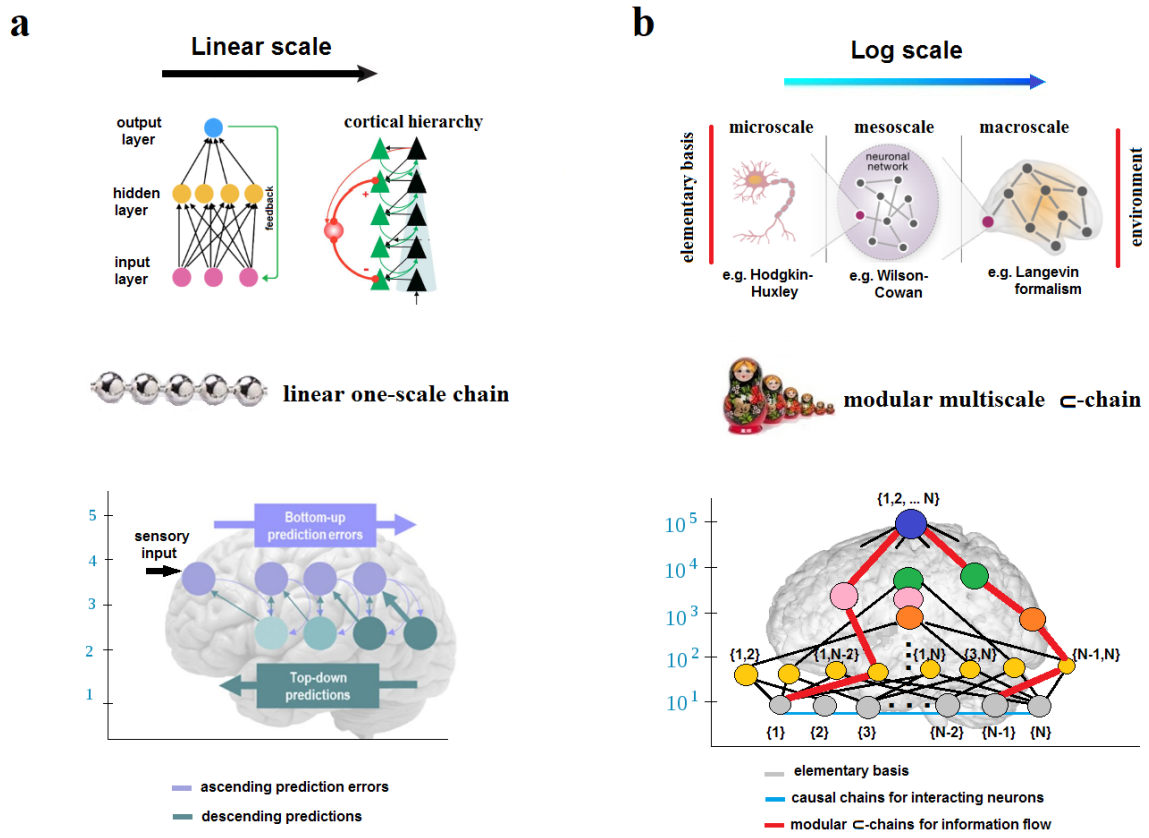


**Figure 9.** (a) Top: This schematic illustrates predictive processing in a flat hierarchy by linear one-scale causal chains, where prediction error (blue circles for superficial pyramidal cells) updates (blue arrows) expectations (teal circles for deep pyramidal cells) at higher layers. Bottom: These posterior expectations then generate predictions of the representations in lower layers via descending predictions (teal arrows). (b) Top: By contrast, the nested brain hierarchy unfolds from individual neurons in the elementary basis to neuronal networks (modules) to the global workspace. Bottom: Causal one-scale chains (blue line) are placed in the elementary basis and provide information flow across the spatial span of the hierarchy via modular ⊂-chains (bold red lines).

Moreover, if the concept of hierarchy is coined to mean multiscale architecture of networks that are nested one within another, even subordinated multilayer structures are not properly hierarchical. The distinction between multilayer (flat) and multiscale (modular) hierarchies becomes especially noticeable in how the two hierarchies can be decomposed by chains. A chain of nodes, all located at the same scale of the flat hierarchy $\mathcal{H}$, is similar to a linear causal chain in the causal set $\mathcal{L}$ (Eqs. 1.1. and 1.2). Conversely, a chain in the modular hierarchy $\Delta$ is spatially extended: each "layer" corresponds to a separate scale, not to its position in the subordinated hierarchy as seen in cortical layers in the brain (Figure 9a). In a modular ⊂-chains, nodes are nested one within another like "matryoshka" dolls (Figure 9b), with symbolic edges representing information flow that can be transmitted across the spatial span of a system. The distinction between these two kinds of chains is important because downward causation is believed to occur along modular ⊂-chains. Therefore, causal analysis must differentiate between flat (subordinated) and modular (spatially nested) hierarchies.

## 9. Causation and Information in Brain Hierarchy

There are now a great number of various theories of consciousness (Evers et al. 2024; Seth and Bayne 2022). Despite discrepancies, many of them converge on the general idea that the brain is an information-processing system, and consciousness is information computed by the brain. In other words, the stream of conscious state, each experienced at a particular moment $t$, is the synergistic

 doi:10.20944/preprints202509.0678.v1



information encoded in the neural structure of the brain at that time. This idea had been encapsulated by Norbert Wiener in his famous statement: "The mechanical brain does not secrete thought "as the liver does bile," as the earlier materialists claimed, nor does it put it out in the form of energy, as the muscle puts out its activity. Information is information, not matter or energy. No materialism which does not admit this can survive at the present day" (Wiener 1962). Another point of convergence among neuroscientists is that conscious experience is a large-scale emergent phenomenon generated by the hierarchical organization of the brain when all modules function as a whole. The modules provide a topological landscape for functionally segregated and differentially integrated neural processing of cognitive and behavioral mechanisms.

The idea that the brain uses hierarchical inference is well-established in neuroscience and provides an explanation for the multilayer anatomical organization of cortical systems (Rao and Ballard 1999). For example, in predictive (Bayesian) processing theory (Friston et al. 2013), neural activity is represented by the cortical hierarchy of ascending prediction errors and descending predictions. In this hierarchy, the sources of forward connections are the superficial pyramidal cell population, and the sources of backward connections are the deep pyramidal cell population (Badcock et al. 2019; Hohwy and Seth 2020). Prediction error is the difference between bottom-up sensory input and top-down predictions of that input. The minimization relies on recurrent neural interactions across different anatomical layers of the cortical hierarchy in which bottom-up signals relay prediction error to higher layers to optimize the posteriors through these feedback mechanisms. Some authors suggest that *multilayer* predictive processing can unfold within a *multiscale* synergistic global workspace to broadcast information to local modules (Safron 2020; VanRullen and Kanai 2021).

The issue of our most interest is this. Can this hierarchically organized global workspace generate a synergistic core to exert downward causation over modular $\subset$-chains? The negative answer follows immediately from the CEP.

The rationale is based on two premises:

1. *Causation can create (and destroy) information, but it cannot move across the spatial span of a modular hierarchy without involving double causation;*
2. *Information can flow across scales and be synergistic* (i.e., *non-additive) in a modular hierarchy, but it cannot generate causation unless neo-Cartesian dualism is covertly admitted.*

At first glance, it may seem that downward causation could still occur via linear chains of subordinated flat hierarchies. Upon closer examination, however, predictive processing is completely based on a flat hierarchy. This means that the terms "bottom-up" and "top-down" can be misleading as they only refer to the anatomical laminar structure of the cortex. All causal chains involved are of the same scale (Figure 9a). Can this type of causation be legitimately labeled as "downward" if the neural network is artificially designed to represent information flow rather than causal chains? Strictly speaking, even information cannot be classified as downward within linear chains of these hierarchies. Although both the flat hierarchy $\mathcal{H}$ and the modular hierarchy $\Delta$ can both be represented by a semilattice, these two are topologically 'orthogonal' to each other.

The causation-information dichotomy can be better understood through the duality of structural and functional connectivity, which is extensively studied in neuroscience. Structural connectivity or the connectome (Sporns et al. 2005; Bennett et al. 2018) refers to direct anatomical links between neurons that give rise to patterns of statistical correlations detected by various neuroimaging techniques, which are then related to functional connectivity (Bullmore and Sporns 2009; Messé et al. 2014; Fukushima et al. 2018). The latter should reflect how neurons contribute to different functional modules involved in perception, cognition, and action. Structural connectivity provides synaptic communication channels for linear causal chains, transmitting information between neurons (Figure 1a), while functional connectivity provides information flow between brain regions. The former constrains the latter, but the reverse is not true as it is encapsulated in the famous dictum "correlation

 doi:10.20944/preprints202509.0678.v1



does not imply causation." This dichotomy becomes more comprehensive in the hierarchical framework.

The multiscale hierarchy $\Delta$ is, by definition, a system composed of modules nested within each other. Modular $\subset$-chains unfold across many spatial scales from the microscale to the macroscale, representing the hierarchy as a whole. Multiple realizability allows for the transmission of information across spatial scales via different $\subset$-chains that reach the same macrostate at the global level (Figure 10a). In causal analysis, multiple realizability is possible due to the causal scope $k$ of individual neurons to bring about neural avalanches via scale transitions (Figure 6a). Thus, not only can a single neuron trigger a behavioral output at the macroscale but initiations caused by different neurons can lead to the same global output.

The multiscale modular organization reflects functional connectivity. Functional connectivity is studied at the mesoscale, while structural connectivity refers exclusively to white-matter fibers at the microscale, where the most significant causal chains determine brain activity. Accordingly, structural connectivity depends on interactions between single neurons in the elementary basis, while functional connectivity results from statistical correlations at the mesoscale of brain activity. Ultimately, conscious states emerge from the consolidated causal chain of brain dynamics at the macroscale, where information from modules becomes synergistically integrated (Figure 10b).

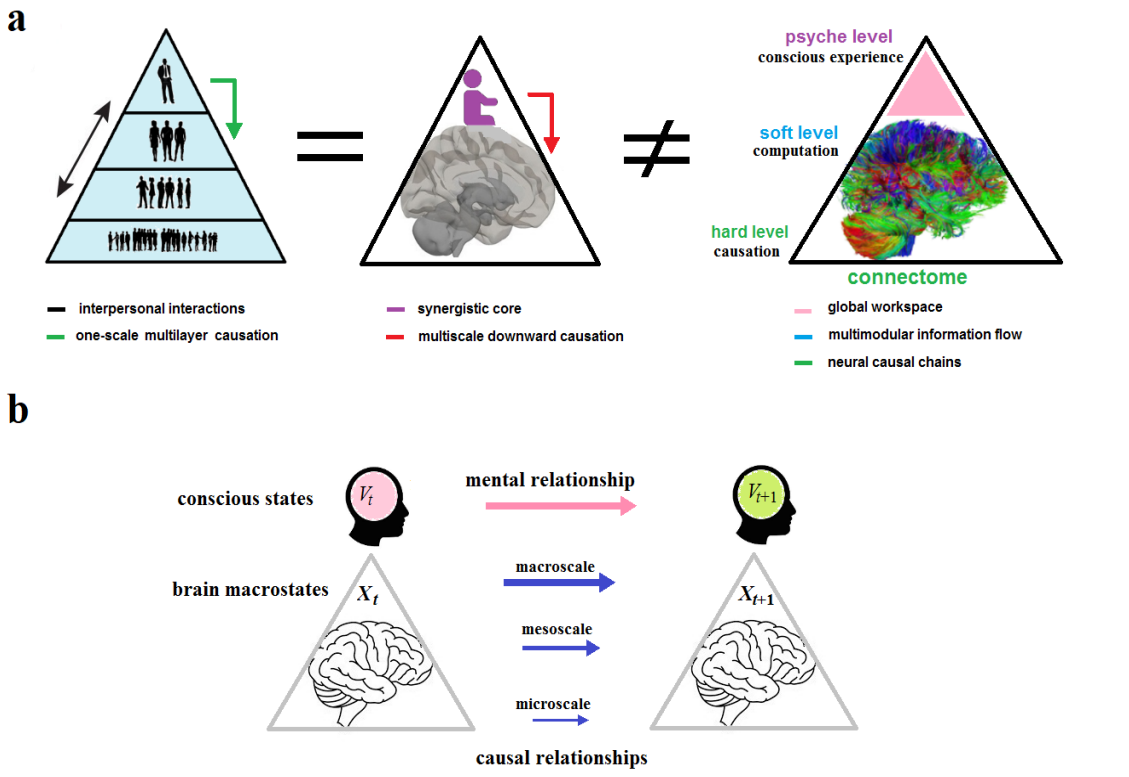


**Figure 10.** (a) In a flat hierarchy such as power hierarchy or pyramidal cells in the cortex, neural causal chains evolve over subordinated layers, not across scales. In a multiscale hierarchy such as the whole brain, downward causation is physically impossible, thus prohibiting the causal emergence. Instead, the connectome is divided into three scale-dependent levels: hard, soft, and psyche levels for neuronal (physical), cognitive (computational), and behavioral (psychic) explanation of emergent conscious experience. (b) According to the CEP, causal chains in brain dynamics are equally driven by each scale separately. Mental cause-like relationships between conscious states provide the coherence of the stream of consciousness exposed to the psyche level by the brain, engaged in Bayesian predictive processing via updating its priors (the generative model of the world) to posteriors based on sensory inputs. These conscious states (information contents) are then presented by supervenient variables $V_t$, which emerge from the corresponding brain macrostates $\boldsymbol{X}_t$ due to synergistic information accumulated in the global workspace.

 doi:10.20944/preprints202509.0678.v1



## 10. Discussion

Downward causation is often discussed in the context of emergent phenomena. In neuroscience, emergence is directly related to the mind-brain problem, whereas downward causation is suggested to account for the causal power of consciousness. Some authors argue that standard reductionist notions of causation in physics are not wrong but simply impoverished (Noble et al. 2019; Ellis and Kopel 2019). In their opinion, a satisfactory understanding of emergent phenomena will ultimately lie in the concepts of downward causation within complex information-processing systems.

Recently, a series of papers have suggested the formal theory of causal emergence, which argues that downward causation can occur due to synergistic effects. The theory is based on the PID formalism (Rosas et al. 2020; Luppi et al. 2021; Mediano et al. 2022), where downward causation is formalized in Equation (8) as an emergent feature of a dynamical system represented by a supervenient variable $V_t$ which has unique predictive power over specific parts of the system that is equated to causal power (Figure 4b). In addition, causal decoupling takes place in Equation (9) when $V_t$ is supposed to have the same causal power over the collective properties of the system itself (Figure 2). Causal emergence is then conceptualized as a symbolic sum of Equations (8) and (9):

causal emergence = causal decoupling + downward causation.

This seems to open a loophole for active consciousness, where conscious states emerging at the macroscale could causally influence brain dynamics driven by neural activity at the microscale. Let us consider this controversial idea in terms of modular hierarchy. Information is physical in origin (Landauer 1961) and can even be equated with the mass-energy equivalence principle (Vopson 2019), but it does not possess its own physical substance to generate the canonical cause-effect relationships beyond those provided by the matter in which it is instantiated. Information physically "resides" in events (represented statistically in state space) and is transmitted by linear causal chains in spacetime, including natural, biological, or manmade channels of communication as a special case. While causation can produce information, the reverse is not true. Information, even if merged in synergy, can flow over many spatial scales via modular $\subset$-chains; however, it does not produce causation.

Granting causal power to consciousness in the above scenario involves not only downward causation, which is dismissed by the CEP, but also neo-Cartesian dualism under the condition that information, placed on an equal footing with matter and energy by Wiener (1962), can have its own causal power in biological systems. Assigning a fictional ontological status to information (Levy 2011; Cardenas-Garcia 2023) is, thus, a prerequisite for mental causation. Remarkably, causal decoupling is compatible with the CEP, which argues that all scales are causally closed (Corollary 1). At the same time, this is precisely why downward causation is forbidden (Corollary 2). Ultimately, the CEP allows preserving the conventional form of free will of consciousness, acting on "behalf" of the brain (Figure 5), but it eliminates the possibility of mental downward causation (Figure 10b).

The theory of causal emergence was in turn inspired by the measure of *effective information*, conditioned on the concept of integrated information $\Phi$, generated by a system of binary elements or logic gates (Hoel et al. 2013). Their approach is based on Pearl's causation, presented in terms of counterfactual interventions. Coarse-graining is assumed to increase effective information, which quantifies how much knowing the system's past reduces uncertainty about its future when all past states are equally likely. Overall, the macro beats the micro by grouping together redundant or noisy elements to increase their cause-effect power. The calculations then show that the information, given by "determinism coefficients" above the maximum entropy distribution, can be higher at the macroscale than at the microscale, as if the former were doing the main causal work within a system.

In fact, the only thing shown there is that a system can appear more ordered "from above" than "from below" because all multi-body circular chains at the microscale become redundant at the macroscale, exhibiting its own consolidated causal structure (Figure 4b). These information-based calculations confirm an intuitively obvious truth about the contextuality of statistical analysis. At each lower (fine-grained) scale of observation, a new set of accessible (counterfactual) states appears, with corresponding probability distributions, altering the orderliness and complexity of the entire system. The consolidation of many linear causal chains at the microscale into a single one-body causal

 doi:10.20944/preprints202509.0678.v1



chain at the macroscale (Figure 8) reduces causal degeneracy of a system. In endogenously coarse-grained systems, the reduction of degeneracy stems from multiple realizability of the same state, the ability of different groups of elements in the elementary basis of a multiscale modular hierarchy to initiate performing the same function or yielding the same output. In this way, information, transmitted by linear causal chains, can flow across scales via modular $\subset$-chains and become synergistic at the global level (Figure 9b).

The main merit of Hoel et al.'s approach is that it captures the nature of autonomy and self-organization that emerge spontaneously in the modular hierarchy of biological systems. No individual (whether a molecule, neuron, bird, or human) within the elementary basis of a multiscale hierarchy is autonomous since their behavior is causally dependent on the behavior of their neighbors. Their dynamics are interdependent, so the multi-body causal chains of such interactions are highly entangled and circular at the microscale. However, at the macroscale, these interactions result in a consolidated one-body causal chain that goes from one state of a system $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$ at time $t$ to its next state at time $t + \Delta t$ (Figure 8). Meanwhile, as follows from the law of conservation of causation, the Markov property is preserved across the spatial span of a system. Now, if $\boldsymbol{X}(t + \Delta t)$ can be predicted from $\boldsymbol{X}(t)$, all the information about microscopic dynamics becomes redundant. Therefore, while a microscopic variable $X_t^1$ has predictive power over its own future state due to the scale-invariance of the Markov property, it cannot predict the evolution of a large number of similar variables in the elementary basis. This property is sometimes referred to as 'horizontality' (Rosas et al. 2018). Thus, in the case of complex self-organizing systems, their "map" can indeed be more informative (but not more causal) than their "territory" which is blind to higher-order phenomena (Hoel 2017). Coarse-graining provides macroscopic variables with more predictive power due to multiple consolidated contributions, making a system, as stated, more ordered from above than from below.

Still, one must distinguish between an epistemological coarse-graining imposed upon a system by observation ad hoc, and an endogenous (ontological) coarse-graining that is intrinsic to the system itself. For example, Barnett and Seth (2023) introduce the concept of *dynamical independence* between microscale and macroscale, conditioned on transfer entropy. Their approach is reminiscent of the notions of information closure or autonomy in (Bertschinger et al. 2008). Dynamical independence is defined in predictive terms: a macroscopic variable is defined to be dynamically-independent if knowledge of the microscopic process adds nothing to prediction of the macroscopic process beyond what the macroscopic process already self-predicts. The authors view dynamical independence as epistemological from a reductionist perspective and conclude that if a macroscopic process appears to emerge as a process in its own right, this apparent autonomy is in the eye of an observer blind to the microscopic dynamics (Barnett and Seth 2023).

Their conclusion, however, cannot be supported by the CEP. According to Corollaries 1, 2, and 3, self-organization and autonomy can emerge at the macroscale of hierarchically organized systems independently of their elementary basis at the microscale, where numerous causal chains are interdependent and highly entangled. The change in the scale of observation does not multiply causation, and downward causation is precluded. Nonetheless, if the change in scale, imposed by observation, is naturally accompanied by endogenous coarse-graining of multiscale hierarchical systems, it reduces a great number of entangled causal chains at the microscale to a consolidated causal chain at the macroscale (Figure 8). This means that large-scale emergent phenomena, firstly, living organisms, are more than just artifacts of macroscopic observations. In particular, the predictive power of macroscopic variables in providing *effective information* for causal emergence by reducing degeneracy at the microscale (Hoel et al. 2013) is a direct consequence of this consolidation, acting like an "information squeezer" (Zhang and Liu 2023), when a system is endogenously coarse-grained.

Finally, the CEP provides an ontological foundation for multilevel selection in evolutionary biology, which raises various philosophical questions, including causal ones (Watson et al. 2022; Okasha 2022). In general, the assumption that natural selection operates not only at the gene level but

 doi:10.20944/preprints202509.0678.v1



also at the levels of organisms and their populations necessarily implies that linear causal chains must be evolutionarily effective at all spatial scales under selection pressure. Yet, since Schrödinger's time, it has been widely accepted that organisms should resist the second law of thermodynamics as the natural enemy of life. From this perspective, the evolution of life is the evolution of biological networks that are able to evade decay to thermodynamical equilibrium (i.e., organic death) on short ontogenetic time scales. Natural selection is the selection of networks that are causally effective for the survival of species over long phylogenetic time scales. There is no upward or downward causation, but selection operates simultaneously at all spatial scales, each exhibiting its own causal structure. Physically, this means that life is necessarily a large-scale phenomenon (starting at the molecular scale due to stable chemical bonds), and all major evolutionary transitions from RNA to human societies (Maynard Smith and Szathmary 1995; Schuster 2016) are driven by scale transitions advancing information integration across hierarchically organized causal structures at each scale.

The CEP (Corollaries 1 and 3) implicitly underlies the renormalizability of dynamical systems, which has been proposed by Vanchurin et al. (2022) as one of the fundamental principles of evolution: Across the entire range of hierarchical organization of evolving systems, a statistical description of faster-changing microscopic variables is feasible through the slower-changing macroscopic variables, making learning valuable at any scale of observation. The authors argue that in a universe without this principle (as opposed to reductionism), it would be impossible for living systems to survive without first discovering fundamental physical laws, whereas complex organisms on our planet have evolved for billions of years before starting to study quantum physics (Vanchurin et al. 2022). In other words, the law of conservation of causation becomes a necessary prerequisite for biological systems to learn, extracting information about the environment from causal relationships at different scales.

The CEP, however, disagrees with the Bayesian models of natural selection (Czégel et al. 2019) if those are ontologically based on Noble's relativity (Noble et al. 2019). Since Bayesian models often equate information flow with causation by associating a reason with a cause, it is not surprising that they admit cross-scale causation, which is forbidden by the CEP. For example, the model of evolutionary synthesis (Friston et al. 2023) assumes that evolution can be described with two random dynamical systems, describing phylogenetic and phenotypic processes coupled over evolutionary timescales via renormalization. The mapping relies on a coarse-graining between slow phylogenetic processes at the population level and fast phenotypic processes at the organism level. Although similar to Vanchurin et al.'s model of evolution as multilevel learning, Friston et al.'s approach generalizes ontogeny (action selection) and phylogeny (model selection) as an interplay between upward and downward causation in cyclical evolutionary processes.

## 11. Conclusions

The CEP is an extension of the relativity postulate in relativity theory. It asserts that not only are all inertial reference frames Lorenz invariant with respect to causal chains by preserving the spacetime interval between two events, but all scales in any reference frame (for any observer) are also causally equivalent. The CEP can be generalized as the law of conservation of causation in terms of the continuity equation. This states that the flow of causation in the universe is conserved across scales.

In causal analysis, the distinction between life and non-life is a matter of multiscale organization. However, even if causal analysis can explain *how* to construct a living system from atoms, the mystery still remains as to *why* life and consciousness would spontaneously emerge from atomic interactions, though, according to the CEP, no "quantity" of causation was added there. While we humans, who are conscious and sentient, capable of learning and creating, are born one day, live our lives, and then inevitably die, nothing of the sort happens at the atomic scale. The atoms we are composed of neither are born nor die. So, which scale tells us the truth about what happens in the universe?

Philosophically, the CEP argues for the stratification of sciences where psychology is not reducible to biology, biology is not reducible to chemistry, and chemistry is not reducible to physics. The classical world emerges from the quantum world at all scales simultaneously, each with its own

 doi:10.20944/preprints202509.0678.v1



causal structure that cannot be derived from the causal structure at another scale. When examining an endogenously coarse-grained system of interest, different scales of observation across its spatial span may offer different causal explanations for its dynamics, but all of them will be valid.

A unified "theory of everything" capable of describing all emergent phenomena across scales at a preferred scale is impossible. Even within biology, multilevel selection can operate consistently at different scales, not reducible to a single scale. It is impossible to explain at the genetic scale why evolution has causally favored one phenotypic trait over another at the organism or population scale. It is an even more unsolvable problem to explain the nature of free will at the atomic scale, i.e., why a conscious being has psychologically chosen one action over another. We might not be able to do so not because of big data involving a computationally intractable number of microevents in observation, but because there is no conscious being at the atomic scale. The universe exists at all spatial scales simultaneously, and we, humans, do not share the same universe with atoms; we inhabit another causal universe.

**Funding:** No funding was received to assist with the preparation of this manuscript.

**Conflict of Interest:** The author has no relevant financial or non-financial interests to disclose.

## References

1. Adami C (1995) Self-organized criticality in living systems. Phys Lett A 203(1):29-32. https://doi.org/10.1016/0375-9601(95)00372-A
2. Allen B, Stacey BC, Bar-Yam Y (2017) Multiscale Information Theory and the Marginal Utility of Information. Entropy 19(6):273. https://doi.org/10.3390/e19060273
3. Badcock PB, Friston KJ, Ramstead MJD (2019) The hierarchically mechanistic mind: a free-energy formulation of the human psyche. Phys Life Rev 31:104-121. DOI: 10.1016/j.plrev.2018.10.002
4. Ballerini M, Cabibbo N, Candelier R, et al. (2008) Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. Proc Natl Acad Sci USA 105:1232-1237. https://doi.org/10.1073/pnas.071143710
5. Barack DL, Miller EK, Moore CI, et al. (2022) A call for more clarity around causality in neuroscience. Trends Neurosci 45(9):654–655. https://doi.org/10.1016/j.tins.2022.06.003
6. Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy are equivalent for Gaussian variables. Phys Rev Lett 103:238701. https://doi.org/10.1103/PhysRevLett.103.238701
7. Barnett L, Seth AK (2023) Dynamical independence: discovering emergent macroscopic processes in complex dynamical systems. Phys Rev E 108:014304. https://doi.org/10.1103/PhysRevE.108.014304
8. Bedau M (1997) Weak emergence. Nous, 31: 375–399. DOI: 10.1111/0029-4624.31.s11.17
9. Bennett SH, Kirby AJ, Finnerty GT (2018) Rewiring the connectome: Evidence and effects. Neurosci Biobehav Rev 88:51–62. https://doi.org/10.1016/j.neubiorev.2018.03.001
10. Bertschinger N, Olbrich E, Ay N, et al. (2008) Autonomy: An information theoretic perspective. Biosystems 91(2):331–345. https://doi.org/10.1016/j.biosystems.2007.05.018
11. Bombelli L, Lee J, Meyer D, et al. (1987) Spacetime as a causal set. Phys Rev Lett 59:521-524. https://doi.org/10.1103/PhysRevLett.59.521
12. Brooks DS, DiFrisco J, Wimsatt WS (2021) Introduction: Levels of Organization: The Architecture of the Scientific Image. MIT Press. https://doi.org/10.7551/mitpress/12389.003.0004
13. Bullmore ET, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–198. https://doi.org/10.1038/nrn2575
14. Cavagna A, Cimarelli A, Giardina I, et al. (2010) Scale-free correlations in starling flocks. Proc Natl Acad Sci USA 107(26): 11865–11870. https://doi.org/10.1073/pnas.1005766107
15. Chang AYC, Biehl M, Yu Y, et al. (2020) Information Closure Theory of Consciousness. Front Psychol 11:1504. https://doi.org/10.3389/fpsyg.2020.01504
16. Corlett PR, Horga G, Fletcher PC, et al. (2019) Hallucinations and Strong Priors. Trends Cogn Sci. 23(2):114–127. DOI: 10.1016/j.tics.2018.12.001

17. Czégel D, Zachar I, Szathmáry E (2019) Multilevel selection as Bayesian inference, major transitions in individuality as structure learning. R. Soc. Open Sci. 6:190202. http://dx.doi.org/10.1098/rsos.190202
18. Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cogn 79:1–37. https://doi.org/10.1016/S0010-0277(00)00123-2
19. Deco G, Kringelbach ML (2017) Hierarchy of information processing in the brain: a novel 'intrinsic ignition' framework. Neuron 94:961–968. https://doi.org/10.1016/j.neuron.2017.03.028
20. Dehaene S, Lau H, Kouider S (2017) What is consciousness and could machines have it? Science 358:486–492. DOI:10.1126/science.aan8871
21. di Santo S, Villegas P, Burioni R, et al. (2018) Landau–Ginzburg theory of cortex dynamics: Scale-free avalanches emerge at the edge of synchronization. Proc Natl Acad Sci USA 115(7):E1356-E1365. https://doi.org/10.1073/pnas.1712989115
22. DiFrisco J, (2019) Kinds of Biological Individuals: Sortals, Projectibility, and Delection. British J Phil Sci 70(3):845–875. DOI: 10.1093/bjps/axy006
23. Evers K, Farisco M, Pennartz CMA (2024) Assessing the commensurability of theories of consciousness: On the usefulness of common denominators in differentiating, integrating and testing hypotheses. Conscious Cogn 119:103668. https://doi.org/10.1016/j.concog.2024.103668
24. Ellis GFR, Kopel J (2019) The Dynamical Emergence of Biology from Physics: Branching Causation via Biomolecules. Front Physiol 9:1966. https://doi.org/10.3389/fphys.2018.01966
25. Ellis GFR (2023) Efficient, Formal, Material, and Final Causes in Biology and Technology. Entropy 25:1301. https://doi.org/10.3390/e2509130
26. Fagerholm ED, Dezhina Z, Moran RJ, et al. 2023. A primer on entropy in neuroscience. Neurosci Biobehav Rev 146:105070. https://doi.org/10.1016/j.neubiorev.2023.105070
27. Farnsworth KD (2025) How Physical Information Underlies Causation and the Emergence of Systems at all Biological Levels. Acta Biotheor 73:6. https://doi.org/10.1007/s10441-025-09495-3
28. Flack JC (2017) Coarse-graining as a downward causation mechanism. Phil Trans R Soc A 375:20160338. https://doi.org/10.1098/rsta.2016.0338
29. Friston K, Schwartenbeck P, FitzGerald T, et al. (2013) The anatomy of choice: active inference and agency. Front Hum Neurosci 7:598. https://doi.org/10.3389/fnhum.2013.00598
30. Friston K, Friedman DA, Constant A, et al. (2023) A Variational Synthesis of Evolutionary and Developmental Dynamics. Entropy 25, 964. https://doi.org/10.3390/e2507096
31. Fukushima M, Betzel R.F, He Y, et al. (2018) Structure-function relationships during segregated and integrated network states of human brain functional connectivity. Brain Struct Funct 223(3):1091–1106. https://doi.org/10.1007/s00429-017-1539-3
32. Cardenas-Garcia JF (2023) Info-Autopoiesis and the Limits of Artificial General Intelligence. Computers 12:102. https://doi.org/10.3390/computers12050102
33. Godfrey-Smith P (2007) Information in biology. Cambridge Companion to Philos Biol 103–119. https://doi.org/10.1017/CCOL9780521851282.006
34. Grasso M, Albantakis L, Lang JP, et al. (2021) Causal reductionism and causal structures. Nat Neurosci 24:1348–1355. https://doi.org/10.1038/s41593-021-00911-8
35. Gutknecht AJ, Wibral M, Makkeh A (2021) Bits and pieces: understanding information decomposition from part-whole relationships and formal logic. Proc R Soc A 477:20210110. https://doi.org/10.1098/rspa.2021.0110
36. Gysi DM, Nowick K (2020) Construction, comparison and evolution of networks in life sciences and other disciplines. J. R. Soc. Interface 17: 20190610. http://dx.doi.org/10.1098/rsif.2019.0610
37. Haken H (1983) Synergetics: An Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology. Springer, Berlin
38. Haken H, Portugali J (2016) Information and Self-organization: A Unifying Approach and Applications. Entropy 18:197. https://doi.org/10.3390/e18060197
39. Hilgetag CC, Goulas A (2020) 'Hierarchy' in the organization of brain networks. Phil Trans R Soc B 375:20190319. https://doi.org/10.1098/rstb.2019.0319

 doi:10.20944/preprints202509.0678.v1




40. Hoel EP, Albantakis L, Tononi G (2013) Quantifying causal emergence shows that macro can beat micro. Proc Natl Acad Sci USA, 110: 19790–19795. https://doi.org/10.1073/pnas.1314922110
41. Hoel EP (2017) When the map is better than the territory. Entropy 19:188. https://doi.org/10.3390/e19050188
42. Hofmeyr JHS (2018) Causation, constructors and codes. BioSystems 164:121–127. https://doi.org/10.1016/j.biosystems.2017.09.008
43. Hohwy J, Seth A (2020) Predictive processing as a systematic basis for identifying the neural correlates of consciousness. Phil Mind Sci 1(II):3. https://doi.org/10.33735/phimisci.2020.II.64
44. Hulswit M (2002) From Cause to Causation: A Peircean Perspective. Springer Dordrecht. https://doi.org/10.1007/978-94-010-0297-4
45. Hunt T, Schooler JW (2019) The easy part of the hard problem: a resonance theory of consciousness. Front. Hum Neurosci, 13:378. https://doi.org/10.3389/fnhum.2019.00378
46. James RG, Barnett N, Crutchfield JP (2016) Information Flows? A Critique of Transfer Entropies. Phys Rev Lett 116:238701. https://doi.org/10.1103/PhysRevLett.116.238701
47. Kaiser M, Hilgetag CC, Kötter R (2010) Hierarchy and dynamics of neural networks. Front Neuroinform 4:112. https://doi.org/10.3389/fninf.2010.00112
48. Kauffman SA (1993) The Origins of Order: Self-Organization and Selection in Evolution. New York: Oxford University Press
49. Kesić S (2024) Universal Complexity Science and Theory of Everything: Challenges and Prospects. Systems 12:29. https://doi.org/10.3390/systems12010029
50. Kim J (2006) Emergence: core ideas and issues. Synthese 151(3):547–559. https://doi.org/10.1007/s11229-006-9025-0
51. Kivelä M, Arenas A, Barthelemy M, et al. (2014) Multilayer networks. J Complex Netw 2:203–271. https://doi.org/10.1093/comnet/cnu016
52. Krakauer D, Bertschinger N, Olbrich E, et al. (2020) The information theory of individuality. Theory Biosci 139:209–223. https://doi.org/10.1007/s12064-020-00313-7
53. Kranke N (2024) Do concepts of individuality account for individuation practices in studies of host–parasite systems? A modeling account of biological individuality. Theory Biosci 143:279–292. https://doi.org/10.1007/s12064-024-00426-3
54. Kwan AC, Dan Y (2012) Dissection of cortical microcircuits by single-neuron stimulation in vivo. Curr Biol 22(16):1459–1467. http://dx.doi.org/10.1016/j.cub.2012.06.007
55. Landauer R (1961) Dissipation and Heat Generation in the Computing Process. IBM J Research Develop 5:183-191. http://dx.doi.org/10.1147/rd.53.0183
56. Levy A (2011) Information in biology: A fictionalist account. Nous 45(4):640–657. https://doi.org/10.1111/j.1468-0068.2010.00792.x
57. Libet B, Gleason CA, Wright EW, et al. (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. Brain 106:623-642. https://doi.org/10.1093/brain/106.3.623
58. Lizier JT, Prokopenko M (2010) Differentiating information transfer and causal effect. Eur Phys J B 73(4): 605–615. https://doi.org/10.1140/epjb/e2010-00034-5
59. Lombardi F, Pepic S, Shriki O, et al. (2021) Quantifying the coexistence of neuronal oscillations and avalanches. arXiv:2108.06686v2
60. London M, Roth A, Beeren L, et al. (2010) Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. Nature 466:123–127. https://doi.org/10.1038/nature09086
61. Luppi AI, Mediano PAM, Rosas FE, et al. (2021) What it is like to be a bit: an integrated information decomposition account of emergent mental phenomena. Neurosci Conscious 2021, niab027. https://doi.org/10.1093/nc/niab027
62. Luppi AI, Mediano PAM, Rosas FE, et al. (2023) A Synergistic Workspace for Human Consciousness Revealed by Integrated Information Decomposition. eLife 12:RP88173. https://doi.org/10.7554/eLife.88173.3

 doi:10.20944/preprints202509.0678.v1




63. Lynn CW, Bassett DS (2019) The physics of brain network structure, function, and control. Nat Rev Phys 1:318–332. https://doi.org/10.1038/s42254-019-0040-8
64. Maynard Smith J, Szathmary E (1995) The major transitions in evolution. Oxford, UK: Oxford University Press.
65. Maynard Smith J (2000) The concept of information in biology. Philos Sci 67:177–194. https://doi.org/10.1017/CBO9780511778759.007
66. Markov NT, Kennedy H (2013) The importance of being hierarchical. Curr Opin Neurobiol 23:187–194. https://doi.org/10.1016/j.conb.2012.12.008
67. Mediano PAM, Rosas FE, Luppi AI, et al. (2022) Greater than the parts: a review of the information decomposition approach to causal emergence. Phil Trans R Soc A 380:20210246. https://doi.org/10.1098/rsta.2021.0246
68. Messé A, Rudrauf D, Benali H, et al. (2014) Relating structure and function in the human brain: relative contributions of anatomy, stationary dynamics, and non-stationarities. PLoS Comput Biol 10:e1003530. https://doi.org/10.1371/journal.pcbi.1003530
69. Meunier D, Lambiotte R, Fornito A, et al. (2009) Hierarchical modularity in human brain functional networks. Front Neuroinform 3:37. https://doi.org/10.3389/neuro.11.037.2009
70. Mihm J, Loch CH, Wilkinson DM, et al. (2010) Hierarchical structure and search in complex organizations. Manage Sci 56:831–848. https://doi.org/10.1287/mnsc.1100.1148
71. Mudrik L, Arie IG, Amir Y, et al. (2022) Free will without consciousness? Trends Cogn Sci 26(7):555-566. https://doi.org/10.1016/j.tics.2022.03.005
72. Noble R, Tasaki K, Noble PJ, et al. (2019) Biological Relativity Requires Circular Causality but Not Symmetry of Causation: So, Where, What and When Are the Boundaries? Front Physiol 10:827. https://doi.org/10.3389/fphys.2019.00827
73. O'Connor T (1994) Emergent properties. American Phil Quart 31:91–104. http://www.jstor.org/stable/20014490
74. Okasha S (2022) The Major Transitions in Evolution—A Philosophy-of-Science Perspective. Front Ecol Evol 10:793824. https://doi.org/10.3389/fevo.2022.793824
75. Pearl J (2000) Causality. Cambridge, UK: Cambridge University Press
76. Piasini E, Panzeri S (2019) Information Theory in Neuroscience. Entropy 21:62. https://doi.org/10.3390/e21010062
77. Rao R, Ballard D (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87. https://doi.org/10.1038/4580
78. Reid AT, Headley DW, Mill RD, et al. (2019) Advancing functional connectivity research from association to causation. Nat Neurosci 22(11):1751–1760. https://doi.org/10.1038/s41593-019-0510-4
79. Rolls ET (2021) Mind Causality: A Computational Neuroscience Approach. Front Comput Neurosci 15:706505. https://doi.org/10.3389/fncom.2021.706505
80. Romero F (2015) Why there isn't inter-level causation in mechanisms. Synthese 192:3731–3755. https://doi.org/10.1007/s11229-015-0718-0
81. Rosas FE, Mediano PAM, Ugarte M, et al. (2018) An Information-Theoretic Approach to Self-Organisation: Emergence of Complex Interdependencies in Coupled Dynamical Systems. Entropy 20:793. https://doi.org/10.3390/e20100793
82. Rosas FE, Mediano PAM, Jensen HJ, et al. (2020) Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. PLoS Comput Biol 16(12): e1008289. https://doi.org/10.1371/journal.pcbi.1008289
83. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52(3):1059-69. https://doi.org/10.1016/j.neuroimage.2009.10.003
84. Safron A (2020) An integrated world modeling theory (IWNT) of consciousness: Combining integrated information and global neuronal workspace theories with the free energy principle and active inference framework; toward solving the hard problem and characterizing agentic causation. Front Art Intell 3:30. | https://doi.org/10.3389/frai.2020.00030

85. Salvaris M, Haggard P (2014) Decoding intention at Sensorimotor timescales. PLOS ONE 9:e85100. https://doi.org/10.1371/journal.pone.0085100
86. Schultze-Kraft M, Birman D, Rusconi M, et al. (2016) The point of no return in vetoing self-initiated movements. Proc Natl Acad Sci USA 113:1080–1085. https://doi.org/10.1073/pnas.1513569112
87. Schurger A, Sitt J. D, Dehaene S (2012) An accumulator model for spontaneous neural activity prior to self-initiated movement. Proc Natl Acad Sci USA 109:E2904–E2913. https://doi.org/10.1073/pnas.1210467109
88. Schuster P (2016) Some mechanistic requirements for major transitions. Phil Trans R Soc B 371:20150439. http://dx.doi.org/10.1098/rstb.2015.0439
89. Seth AK, Bayne T (2022) Theories of consciousness. Nat Rev Neurosci 23:439–452. https://doi.org/10.1038/s41583-022-00587-4
90. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27: 623–656.
91. Signorelli CM, Boils JD, Tagliazucchi E, et al. (2022) From brain-body function to conscious interactions. Neurosci Biobehav Rev 141:104833. https://doi.org/10.1016/j.neubiorev.2022.104833
92. Simon H (1969) The Sciences of the Artificial. MIT Press, Cambridge, Massachusetts
93. Sorkin RD (2009) Light, links and causal sets. J Phys Conf Ser 174:012018. DOI: 10.1088/1742-6596/174/1/012018
94. Sperry RW (1980) Mind-brain interaction: mentalism, yes; dualism, no. Neurosci 5:195–206. https://doi.org/10.1016/0306-4522(80)90098-6
95. Sporns O, Tononi G, Kötter R (2005) The human connectome: a structural description of the human brain. PLoS Comput Biol 1:e42. https://doi.org/10.1371/journal.pcbi.0010042
96. Tehrani-Saleh A, Adami C (2018) Can Transfer Entropy Infer Information Flow in Neuronal Circuits for Cognitive Processing? Entropy 22(4):385. https://doi.org/10.3390/e22040385
97. Timme NM, Lapish C (2018) A tutorial for information theory in neuroscience. eNeuro 5(3). https://doi.org/10.1523/ENEURO.0052-18.2018
98. Triggiani AI, Kreiman G, Lewis S, et al. (2023) What is the intention to move and when does it occur? Neurosci Biobehav Rev 151:105199. https://doi.org/10.1016/j.neubiorev.2023.105199
99. Tononi G, Albantakis L, Boly M, et al. (2022) Only what exists can cause: An intrinsic view of free will. https://arxiv.org/abs/2206.02069
100. Turkheimer FE, Rosas FE, Dipasquale O, et al. (2022) A Complex Systems Perspective on Neuroimaging Studies of Behavior and Its Disorders. The Neuroscientist 28(4):382–399. https://doi.org/10.1177/10738584219947
101. Turkheimer FE, Hellyer P, Kehagia AA, et al. (2019) Conflicting Emergences. Weak vs. strong emergence for the modelling of brain function. Neurosci Biobehav Rev 99:3–10. https://doi.org/10.1016/j.neubiorev.2019.01.023
102. Ursino M, Ricci G, Magosso E (2020) Transfer Entropy as a Measure of Brain Connectivity: A Critical Analysis with the Help of Neural Mass Models. Front Comput Neurosci 14:45. https://doi.org/10.3389/fncom.2020.00045
103. Vanchurin V, Wolf YI, Katsnelson MI, et al. (2022) Toward a theory of evolution as multilevel learning. Proc Natl Acad Sci USA 119:e2120042119. https://doi.org/10.1073/pnas.2120037119
104. VanRullen R, Kanai R (2021) Deep learning and the global workspace theory. Trends Neurosci 44:692–704. https://doi.org/10.1016/j.tins.2021.04.005
105. Varley TF, Hoel E (2022) Emergence as the conversion of information: a unifying theory. Phil Trans R Soc A 380:20210150. https://doi.org/10.1098/rsta.2021.0150
106. Vohryzek J, Cabral J, Vuust P, et al. (2022) Understanding brain states across spacetime informed by whole-brain modelling. Phil Trans R Soc A 380:20210247. https://doi.org/10.1098/rsta.2021.0247
107. Vopson MM (2019) The mass-energy-information equivalence principle. AIP Advance 9:095206. http://doi.org/10.1063/1.5123794
108. Walker SI, Davies PCW (2013) The algorithmic origins of life. J R Soc Interface 10:20120869. http://dx.doi.org/10.1098/rsif.2012.0869

109. Walker SI, Davies PCW (2017) The "Hard Problem" of Life. In: Walker SI, Davies PCW, Ellis GFR (eds) From Matter to Life: Information and Causality. Cambridge University Press, pp 19-37. https://doi.org/10.1017/9781316584200.002
110. Watson RA, Levin M and Buckley CL (2022) Design for an Individual: Connectionist Approaches
111. to the Evolutionary Transitions in Individuality. Front Ecol Evol 10:823588. http://doi.org/10.3389/fevo.2022.823588
112. Watts DJ, Strogatz S (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442. https://doi.org/10.1038/30918
113. Weichwald S, Peters J (2021) Causality in cognitive neuroscience: concepts, challenges, and distributional robustness. J Cogn. Neurosci 33(2):226–247. https://doi.org/10.1162/jocn_a_01623
114. Wibral M, Lizier JT, Priesemann V (2015) Bits from brains for biologically inspired computing. Front Robot AI 2(5):1-25. https://doi.org/10.3389/frobt.2015.00005
115. Wiener N (1962). Cybernetics: or Control and Communication in the Animal and the Machine. Cambridge MA: MIT Press
116. Williams PL, Beer RD (2010) Nonnegative decomposition of multivariate information. https://doi.org/10.48550/arXiv.1004.2515
117. Woodward J (2003) Making Things Happen. A theory of Causal Explanation. Oxford, MI, USA: Oxford University Press
118. Yuan B, Zhang J, Lyu A, et al. (2024) Emergence and Causality in Complex Systems: A Survey of Causal Emergence and Related Quantitative Studies. Entropy 26:108. https://doi.org/10.3390/e26020108
119. Yurchenko SB (2022) From the origins to the stream of consciousness and its neural correlates. Front Integr Neurosci 16:928978. https://doi.org/10.3389/fnint.2022.928978
120. Yurchenko SB (2023a) Is information the other face of causation in biological systems? BioSystems 229:104925. https://doi.org/10.1016/j.biosystems.2023.104925
121. Yurchenko SB (2023b) A systematic approach to brain dynamics: cognitive evolution theory of consciousness. Cogn Neurodyn 17:575–603. https://doi.org/10.1007/s11571-022-09863-6
122. Yurchenko SB (2024) Panpsychism and dualism in the science of consciousness. Neurosci Biobehav Rev 165:105845. https://doi.org/10.1016/j.neubiorev.2024.105845
123. Yurchenko SB (2025) On a physical theory of causation in multiscale analysis of biological systems. *Authorea.* DOI: 10.22541/au.175458667.73208046/v1
124. Zhang J, Liu K (2023) Neural Information Squeezer for Causal Emergence. Entropy 25:26. https://doi.org/10.3390/e25010026