The Causal Core

A centralist principle of causality from the prokaryotic cell to self, ego, and agency

Abstract

This article develops the thesis that living systems form a centralist internal asymmetry from their very first appearance, a structure I refer to as the causal core. This core is not a morphological compartment and not a localisable site. It is a functional attractor that emerges from superposed thermodynamic flows within an operationally closed space, constitutes itself in the prokaryotic cell, and recurs in formally invariant fashion at every higher organisational level of the living. The article draws together the empirical evidence for this continuity from five domains, discusses five possible objections, and finally applies the model to self, ego, and agency. These three terms are not reconstructed as independent cognitive modules but as three descriptive directions of the same causal core at the highest level of biological complexity.

1. Starting hypothesis

The basic idea can be developed in three steps. First, living systems come into being when autocatalytic reaction cycles are enclosed by a membrane. This enclosure establishes operational closure in the sense of Maturana and Varela. Second, within the resulting interior, thermodynamic flows arise that pass through the membrane in exchange with the environment and that superpose within the interior. Third, this superposition necessarily produces an internal asymmetry, because the flows are not homogeneously distributed but directed and variable in density. A zone of maximal reaction density forms, organising the system dynamics as a functional attractor. This zone is the causal core.

Crucially, the causal core is not a place in the geometrical sense but a topological condensation of causal integration. It can concentrate locally, as at the poles of Caulobacter crescentus; it can oscillate, as in the Min proteins of Escherichia coli; or it can be distributed, as in the hub regions of the human connectome. What identifies these as manifestations of the same principle is not spatial location but formal structure: flow-driven symmetry breaking in an operationally closed space that establishes a functional attractor.

The main thesis of the article is that this principle recurs at every higher organisational level of the living, not as an analogical metaphor but as a formal invariance under varying biological realisations. Moreover, the principle has not only been functionally conserved during evolution but also morphologically manifested, first in the nucleus of the eukaryotic cell, later in the centralised architecture of the nervous system, and finally in the brain of vertebrates. It is therefore not a contingent arrangement that each generation would have to reinvent, but a genetically anchored constructional principle, comparable to the

homochirality of biological molecules (Frank 1953, Kondepudi and Asakura 2001). Just as the once established L-form of amino acids and D-form of sugars became, through autocatalytic amplification, a universal trait of the biosphere, the centralist internal asymmetry, once stabilised by flow-driven symmetry breaking in the prokaryotic cell, has shaped the entire subsequent evolutionary path.

A conceptual clarification is required at the outset. The causal core is grounded thermodynamically throughout what follows, that is, in terms of non-equilibrium flows and entropy production. This does not exclude information-theoretic descriptions but situates them. In the view defended here, information is not an autonomous level of reality alongside the thermodynamic level, but a descriptive language of the same flows. The causal core is ontologically constituted by thermodynamics and can be described informationally, without any causal interaction between the two levels. This ontologically monist, semantically pluralist position is relevant to what follows because it prevents integration, representation, or information from being invoked as independent causal factors alongside flow dynamics.

The article develops this thesis in three steps. Section 2 reconstructs the empirical chain of evidence from prokaryotic polarity to the integration dynamics of the brain. Section 3 discusses five objections and defends the thesis in a methodologically refined form. Section 4 applies the model to self, ego, and agency.

2. The empirical chain of evidence

2.1 Flow asymmetry in the prokaryotic cell

Prokaryotes have no internal membranes, no organelles, and no morphological nucleus. Standard textbook accounts suggest that they are therefore structurally simple and homogeneously organised. Research over the past two decades has thoroughly revised this picture. Prokaryotes exhibit a pronounced functional internal asymmetry that does without compartmentalisation and yet realises a centralist causal structure.

Work by Lucy Shapiro and Christine Jacobs-Wagner on Caulobacter crescentus has shown that specific proteins such as PopZ, PleC, and DivJ accumulate at defined cell poles and function there as organisational centres for replication, division, and signal processing (Shapiro, McAdams, and Losick 2009, Jacobs-Wagner 2004, Ebersbach and Jacobs-Wagner 2007). This polarity is not architecturally predetermined but emerges from reaction-diffusion dynamics and flow asymmetries within the enclosed cellular space. Two observations are especially relevant for the thesis advanced here. First, polarity is spontaneously re-established in both daughter cells after division, which indicates that it is not inherited but repeatedly generated from the thermodynamic conditions of the enclosed space. Second, polarity collapses when the energy supply is interrupted, confirming the flow dependence of the structure.

The principle appears even more cleanly in the Min-protein oscillations of Escherichia coli. Petra Schwille and Erwin Frey have reconstructed in vitro how purely thermodynamically driven non-equilibrium flows, specifically ATP hydrolysis combined with membrane binding,

spontaneously generate spatial polarity patterns that determine the site of cell division (Loose, Fischer-Friedrich, Ries, Kruse, and Schwille 2008, Halatek and Frey 2018). The Min oscillation is a paradigm case of spontaneous symmetry breaking within a membrane-enclosed space that requires no prior structural information. Halatek and Frey have shown mathematically that such systems belong to a generic class of reaction-diffusion equations whose solutions form centralist polarity patterns across wide parameter regimes. This constitutes formal proof that the causal core is not a contingent biological peculiarity but a necessary consequence of certain classes of thermodynamic systems.

Complementary to these findings are the studies of biomolecular condensates and liquid-liquid phase separation by Anthony Hyman, Clifford Brangwynne, and colleagues (Brangwynne et al. 2009, Hyman, Weber, and Jülicher 2014, Banani et al. 2017). Membraneless density zones form within cells and function as reaction centres. These zones provide perhaps the clearest empirical example of a central causal structure without structural boundary. They are neither organelles nor mere concentration differences but dynamically stabilised phases, maintained by thermodynamic flows and locally centralising specific biochemical processes.

The general thermodynamic framework for all these observations was formulated already in the 1970s by Ilya Prigogine and his school. The theory of dissipative structures shows that non-equilibrium flows in enclosed spaces spontaneously generate symmetry-breaking centres once the distance from thermodynamic equilibrium exceeds a critical threshold (Prigogine and Nicolis 1977, Kondepudi and Prigogine 1998). In this sense the causal core is not a biological oddity but a thermodynamic necessity under the conditions of membrane-enclosed autocatalysis. Biology did not invent this principle, it inherited it.

2.2 Morphological manifestation in the eukaryotic nucleus

The transition from the prokaryotic to the eukaryotic cell marks the first great step in which the previously purely flow-dynamic causal core acquires a morphological form. Work by Nick Lane and William Martin has reconstructed this transition energetically and shown that the centralisation of information processing in the nucleus goes hand in hand with the outsourcing of bioenergetics to the mitochondria (Lane and Martin 2010, Lane 2015). What is decisive for the thesis advanced here is that the nucleus does not arise as an afterthought to an otherwise finished cell but appears as the spatial condensation of a functional centrality that was already operative. The eukaryotic morphology structurally consummates what was prokaryotically established in dynamic terms.

This reading is not trivial and must at the same time be guarded against a misunderstanding. The claim of continuity of the centralist principle does not imply that eukaryogenesis is not a biological innovation. The spatial separation of transcription and translation, introns and alternative splicing, the regulatory complexity of gene expression, and the capacity to form differentiated tissues are consequential novelties that became possible only through the membrane-enclosed nucleus. The perspective proposed here does not underestimate this innovation but situates it. The morphologisation of the causal core in the nucleus achieves two things at once. It stabilises an older principle at a new level, and through this stabilisation it opens up new functional degrees of freedom that were not available in the

purely dynamic realisation of prokaryotic polarity. Continuity of principle and innovation of realisation are not mutually exclusive but mutually conditioning. A principle can be innovatively realised anew only where it has already been operative.

2.3 Polarity of the nerve cell

The nerve cell repeats the principle at its own level. Every neuron is structurally and functionally profoundly asymmetric. The dendritic input pole, the somatic integration centre, and the axonal output pole form a directed processing axis maintained by polarly distributed receptors, asymmetric ion flows, and active transport along the microtubule cytoskeleton. Work by Carlos Dotti, Gary Banker, and later Frank Polleux on neuronal polarity formation shows that this asymmetry arises in the growth cone from reaction-diffusion dynamics that formally realise the same type of flow-driven symmetry breaking as prokaryotic polarity (Dotti, Sullivan, and Banker 1988, Polleux and Snider 2010, Arimura and Kaibuchi 2007).

Notably, the molecular players involved, especially small GTPases, kinases, and cytoskeletal components, belong to gene families whose deep evolutionary roots reach back into prokaryotic and early eukaryotic precursors. These are not strict one-to-one homologies in the classical sense but a conserved polarity machinery that has been functionally reinterpreted and exapted in different contexts (Wedlich-Söldner and Li 2003, Goehring and Grill 2013). The continuity lies not in the identical function of homologous molecules but in the evolutionary availability of a class of flow-responsive protein systems that realise the same formal pattern at different levels of organisation. This grounds the continuity thesis molecularly and materially without requiring a strong homology claim.

2.4 Laminar asymmetry of the cortex

At the mesoscopic level the principle appears in the laminar architecture of the neocortex. Work by Rodolfo Llinás and Stewart Shipp on feedforward-feedback asymmetry between cortical layers shows that ascending and descending flows between areas are not symmetrically exchangeable (Shipp 2007, Bastos et al. 2012). Granular layers preferentially receive ascending inputs, while supragranular and infragranular layers send descending projections back. This laminar asymmetry is not a descriptive convention but a causal directedness of information flow that constitutes the cortex as a directed integration organ. The hierarchy of cortical areas expresses, at the tissue level, a centralist principle in which higher integration levels figure as centres relative to peripheral processing regions.

2.5 Core-periphery asymmetry of the connectome

At the network level of the whole brain, Martijn van den Heuvel and Olaf Sporns have shown that the human connectome forms a so-called rich club, a small group of about twelve hub regions (precuneus, superior frontal and parietal cortex, thalamus, hippocampus, putamen) that are exceptionally well interconnected and carry a disproportionately large share of the functional integration of the entire brain (van den Heuvel and Sporns 2011, van den Heuvel et al. 2012). Information flow is not evenly distributed but converges on a topologically privileged core.

What is decisive for the continuity thesis is that this core does not exist as a morphologically delimited organ. There is no bounded central organ of the brain, only a zone of maximal

causal integration density. This zone is a density zone within a flow, topologically analogous to the prokaryotic causal core. Lesions of individual hub regions produce characteristic disturbances of global integration that cannot be reduced to local deficits (Crossley et al. 2014). A methodological clarification is in order at this point, because talk of formal invariance between the Min oscillation and the rich club might appear to conflate heterogeneous phenomena. The Min oscillation is a dynamic symmetry breaking in real time; the rich club is a statistical property of a structural connectivity matrix. The bridge between the two runs through the insight that structural connectivity is nothing but sedimented dynamics. The hub topology of the connectome does not arise independently of flows but is shaped during development by activity-dependent synaptogenesis, Hebbian plasticity, and firing correlations (Katz and Shatz 1996, Petersen and Sporns 2015, Betti and Hagmann 2023). Nodes with the highest firing rates and strongest synchronous activity become hubs over the course of maturation. Topology, in this sense, is frozen flow history. Conversely, the location of the Min oscillation is not independent of structure but follows from cell geometry, which itself emerged from earlier polarity cycles. The distinction between dynamics and topology is therefore a matter of observational timescale, not an ontological divide. On long timescales, dynamics become structure; on short timescales, structure is the arena of dynamics. The formal invariance of the causal core operates precisely at this interface.

2.6 Thermodynamic asymmetry of brain dynamics

Particularly pertinent to the model proposed here is the work of Morten Kringelbach, Gustavo Deco, and their collaborators on brain dynamics as a driven non-equilibrium process (Deco et al. 2019, Kringelbach and Deco 2020). Methods of statistical physics show that brain activity in the waking state exhibits the features of a dissipative system whose detailed balance is broken. There are directed flows, measurable entropy production, and an asymmetry of the time direction.

The most precise empirical confirmation to date is provided by Christopher Lynn, Dale Zhou, Danielle Bassett, and colleagues in their 2021 article in the Proceedings of the National Academy of Sciences entitled "Broken detailed balance and entropy production in the human brain" (Lynn et al. 2021). They show that the asymmetry of neural transitions is strongest during cognitively demanding states and declines during sleep and under anaesthesia. This is empirically the signature that the continuity thesis predicts. The brain is a thermodynamically driven system that stabilises an internal asymmetry, and this asymmetry is not an epiphenomenal accompaniment of activity but its constitutive form. The causal core of prokaryotes and the broken detailed balance of the waking brain are expressions of the same principle at different levels of complexity.

2.7 Ontogeny and plasticity

The presentation so far has been phylogenetic and synchronic. It shows that the same principle operates at every organisational level but does not clarify how the causal core comes about ontogenetically in each individual. This gap must be closed, lest it leave the impression that genetic anchoring alone produces the core. In fact, the core arises anew in every development, through the same dynamics of flow-driven symmetry breaking that also characterise its phylogenetic establishment.

Activity-dependent network maturation, as described by Katz and Shatz (1996), shows that neural connections are not fixed by a finished genetic programme but shaped by spontaneous and later stimulus-driven activity. Firing correlations between neurons strengthen connections, while uncorrelated activity weakens them. In real time this Hebbian dynamic is a flow phenomenon; on developmental timescales it sediments into hub topology. Work by Hagmann and colleagues on connectome development in childhood (Hagmann et al. 2010) confirms that the rich club is not fully genetically prewired but emerges through maturation processes between birth and late adolescence. The genetic base architecture provides the scaffold; the actual centralisation emerges from the individual history of connectivity.

Hensch and colleagues have further shown that the malleability of these integration networks is bound to critical periods of particularly high plasticity, during which flow-driven symmetry breakings fix topology durably (Hensch 2005). Once these periods close, the structure becomes more stable but not immutable. Even in the adult brain, hub topology remains dynamically modulated, and pathological states can shift local integration density.

From this an important refinement of the continuity thesis follows. The causal core is neither exclusively genetically determined nor exclusively ontogenetically acquired. It is the result of an interaction between a genetically provided polarity machinery and an individually realised history of flows. This structure is formally identical to the emergence of prokaryotic polarity, which likewise arises from the interplay of a conserved molecular endowment and local flow conditions. Ontogeny thus recapitulates on the individual timescale what phylogeny accomplished on the evolutionary timescale. The causal core is not a product but a process that enacts itself anew in every living system.

3. Objections and responses

A continuity thesis reaching from the prokaryotic cell to the human brain must defend itself against several possible objections. The five most weighty are discussed in what follows.

3.1 The charge of mere analogy

The first and perhaps most important objection is that talk of the same principle confuses formal similarity with genetic or causal continuity. Symmetry breaking in Min oscillations and broken detailed balance in brain dynamics are both describable as non-equilibrium processes, yet this does not establish that they realise the same principle. Similar formulas describe quite different phenomena in physics.

The response must show that continuity is not merely formal but evolutionary. The principle of centralist internal asymmetry arose in the prokaryotic cell and has been conserved in every more complex structure since, not because symmetry breaking has been repeatedly reinvented but because the polarity machinery once established, namely reaction-diffusion systems with membrane-bound ATP hydrolysis and polarly distributed GTPases, has been carried over in all descendants. The molecular homology of the components involved, from prokaryotes to neurons, is empirically documented. Continuity is therefore not merely formal but material.

3.2 The intermediate steps

The second objection concerns the intermediate steps. More than two billion years of evolution lie between prokaryotic polarity and neuronal polarity. The thesis must not skip over these intermediate steps. Work on cell polarity in eukaryotic cells (Wedlich-Söldner and Li 2003, Goehring and Grill 2013) shows that the same flow-driven symmetry breaking can be demonstrated from prokaryotes through yeasts, amoebae, and plant cells to metazoan cells. The intermediate steps are empirically documented; continuity is not postulated but measured. Particularly striking is the polarisation of the Cdc42 system in Saccharomyces cerevisiae, which mathematically exhibits the same bifurcation type as the Min oscillation and likewise arises from purely flow-driven symmetry breaking.

3.3 The charge of structuralism

The third objection is the classical charge of structuralism. If symmetry breaking appears at every level, the principle threatens to become explanatorily empty. Turbulence, convection cells, and snowflakes also break symmetries, yet they do not form a causal core. The response lies in the specificity of the claim. Not every symmetry breaking counts, only the centralist kind, that is, the one that forms a functional attractor with integration density. Decisive is the combination of operational closure, autocatalytic self-amplification, and flow integration. Turbulence meets none of these three conditions; convection cells meet them only partially. The causal core is a specific class of symmetry breaking, namely that of autocatalytically closed systems, which establishes self-reference and draws integration upon itself.

3.4 Relation to autopoiesis

The fourth objection comes from systems theory. Maturana and Varela have already formulated a continuity thesis with autopoiesis. What does the causal core add? The answer lies in the supplement of an internal structure. Autopoiesis describes operational closure as the achievement of the membrane, that is, the demarcation of an inside from an outside. What it does not describe is the internal differentiation of this inside. Only within the already closed space do the superposed thermodynamic flows generate a centralist asymmetry that inscribes a causal core into operational closure. The model proposed here therefore does not replace autopoiesis but extends it. It shows how, from the flow dynamics made possible by closure, a functional attractor arises that at higher organisational levels comes to be morphologically manifested in the nucleus and in the brain.

3.5 The methodological notion of identity

The fifth objection is methodological. Speaking of the same principle presupposes a relation of identity that is rarely cleanly sustained in biology. Evolution works with co-option, exaptation, and convergence rather than with principle conservation in the strict sense. The continuity thesis should therefore be formulated not as a claim of identity but as a claim of formal invariance under varying biological realisations. This weaker formulation is empirically defensible and philosophically precise. It does not assert that prokaryotic polarity and brain dynamics are the same thing but that they fall under the same formal description and are connected by a conserved molecular machinery.

4. Formal operationalisation

The causal core has so far been introduced as a topological-functional attractor. This raises the methodological question of its formal measurability and empirical falsifiability. The position developed here is that the causal core is not to be identified with any single mathematical formalism but becomes empirically accessible through several established formalisms, each of which quantifies a different aspect of the same phenomenon.

First, transfer entropy and the theory of directed information flows, as developed by Schreiber (2000) and Lizier (2013), allow a direct measurement of asymmetry between network nodes. Unlike symmetric correlation measures, transfer entropy captures directed dependence, that is, the fact that one node predictively influences another. The broken detailed balance empirically demonstrated by Lynn et al. 2021 can be formally described as non-vanishing net transfer entropy between macrostates. In this language the causal core manifests itself as the zone of maximal incoming directed information, measured by integrated transfer entropy across a node.

Second, network control theory, developed by Gu et al. (2015) and Pasqualetti, Zampieri, and Bullo (2014), quantifies how strongly individual nodes can steer the dynamics of a network. Average controllability measures the capacity of a node to bring the whole system into easily reachable states; modal controllability measures its capacity to induce hard-to-reach states. Empirical work shows that the hub regions of the rich club are precisely those nodes whose controllability is disproportionately high. The causal core is thus operationalised not only topologically but dynamically.

Third, stochastic thermodynamics as formulated by Seifert (2012) makes it possible to calculate the thermodynamic costs of integration explicitly. Entropy production, free-energy rate, and heat dissipation can be reconstructed for biological networks from time series. For the causal core one would predict that nodes of maximal integration density also exhibit maximal local entropy production. This prediction is empirically testable and builds a direct bridge from the Prigogine formalism to neural network dynamics.

Important for the methodological position is that the causal core is designated as the phenomenon measurable through these formalisms, not identified with any one of them. The three approaches quantify different aspects (directed information flow, dynamical control, thermodynamic cost) that converge at the integration centre. The empirical prediction of the continuity thesis is that these three quantities, at every organisational level investigated from the Min oscillation to the human connectome, culminate at the same topological structures. This prediction is testable and therefore falsifiable.

Concrete experimental predictions can be derived from the model that other theories do not generate in this form. First, the prognosis that a local suppression of ATP hydrolysis in hub regions of the connectome, for instance by optogenetic or pharmacological intervention, should not only produce local loss of activity but also collapse the topological identity of the core, measurable as homogenisation of network statistics and reduction of directed information flows. Second, the prognosis that altered states of consciousness such as anaesthesia, deep sleep, or psilocybin-induced states should be accompanied by a specific

reduction of broken detailed balance in hub regions, while peripheral regions may preserve it. Third, the prognosis that disturbances of self, ego, and agency in clinical populations should correlate with a specific reduction of transfer entropy between default mode network nodes that exceeds the general reduction of connectivity. These predictions are experimentally addressable and bridge the theoretical reconstruction with empirical research.

5. Theoretical positioning

The model proposed here stands within a broader landscape of theoretical approaches that thematise integration, self-organisation, and cognition. A brief positioning sharpens the distinctiveness of the view and makes its affinities visible.

Integrated Information Theory (Tononi 2008, Tononi et al. 2016) is the most obvious interlocutor because it likewise posits integration as a central principle. Its measure Phi quantifies the irreducibility of a system to independent parts and formally corresponds to an aspect of what is here called the causal core. The difference lies in grounding. IIT postulates integration axiomatically as a fundamental phenomenal property, whereas the causal core is derived from the thermodynamic genesis of the living. The position defended here therefore avoids the panpsychism characteristic of IIT, because integration is not attributed to every physical structure but only to those that arise through flow-driven symmetry breaking in operationally closed systems.

Global Workspace Theory (Baars 1988, Dehaene and Changeux 2011) describes functionally what topologically corresponds to the rich club, namely a central integration platform on which different processing modules share their results. GWT is therefore a functional description of the surface of the causal core without addressing its thermodynamic genesis. The model proposed here complements GWT by asking why a global workspace forms in the first place and answering with the evolutionary conservation of centralist internal asymmetry.

The Predictive Processing framework and the Free Energy Principle after Friston (2010) occupy a special position. Free energy is originally a thermodynamic quantity but is interpreted by Friston in information-theoretic terms as a measure of surprise. This yields a formal bridge that simultaneously marks a difference. The position developed here treats thermodynamic and information-theoretic description as two languages of the same flows without positing any causal interaction between them. This avoids the category errors to which a strong informational reading of the Free Energy Principle is susceptible.

Enactive cognitive science in the tradition of Varela, Thompson, and Di Paolo (Thompson 2007, Di Paolo, Buhrmann, and Barandiaran 2017) is the closest relative of the position developed here, since it treats biological autonomy, autopoiesis, and cognition as a continuum. The difference lies in emphasis. Enactive approaches foreground embodiment and situatedness of cognitive processes, whereas the model proposed here foregrounds thermodynamic internal structure. The two perspectives are complementary. The causal core

describes internal centralisation; enactivism describes external coupling. A fully developed theory would need to integrate both.

Finally, the model is to be distinguished from emergentist and panpsychist approaches. It does not claim new properties at higher organisational levels but the recurrence of the same formal principle in different realisations. And it does not attribute integration to every physical structure but only to those that arise through operationally closed autocatalysis and flow-driven symmetry breaking. This restriction makes the position empirically substantive and philosophically parsimonious.

A final clarification concerning the philosophical scope is in order. The article develops the structural and functional foundation of self, ego, and agency. The phenomenal, subjective, and sociocultural dimensions of these phenomena are deliberately set aside. This is not because these dimensions are unimportant but because their reconstruction requires its own line of argument. The position developed here describes the structural precondition on which phenomenal shaping and cultural formation rest, without explaining these themselves. A comprehensive theory of the self would have to connect the thermodynamic foundation sketched here with a theory of intersubjectivity, symbolic mediation, and bodily situatedness.

6. Application to self, ego, and agency

The productive upshot of the continuity thesis shows itself in its application to self, ego, and agency. The classical intuition treats these three terms as self-standing phenomena or even as mutually independent cognitive modules. The perspective developed here suggests a different reading. Self, ego, and agency do not designate three things but three descriptive directions of the same causal core at the highest level of complexity. The self describes it as internal structure, the ego describes it as self-reference, agency describes it as outward action.

6.1 The self as internal structure

Bjorn Merker compiled the evidence in Behavioral and Brain Sciences in 2007 that children with hydranencephaly, who largely lack the cerebral hemispheres, nonetheless show clear signs of affectivity, preference formation, and goal-directed interaction (Merker 2007). The minimal self is not anchored cortically but in subcortical integration structures of the brainstem, in particular the periaqueductal grey, the superior colliculus, and the parabrachial nucleus. Antonio Damasio has integrated this finding into the theory of a proto-self whose basis is not representational but consists in basal integration density (Damasio 1999, Damasio 2010).

From the perspective of the causal core, the self is precisely the stable internal asymmetry that distinguishes a system from its environment and integrates it within itself. It is not a content of the system but its central structure. This explains why the self can be preserved under severe cortical damage provided that the subcortical integration cores remain intact. The self is not bound to the highest processing level but to whichever integration zone is densest. This reading also makes it intelligible why self-related states vary in intensity during meditation, trance, or dissociation without the person losing their identity. The self is not a content that is present or absent but an integration density that rises and falls.

6.2 The ego as reflexive signature

At the cortical level, work by Marcus Raichle and Kalina Christoff on the default mode network shows that self-related processing depends on the integration density of a network whose centre is formed by the precuneus and the medial prefrontal cortex (Raichle 2015, Christoff et al. 2016). This network is topologically congruent with the rich club of the connectome. Georg Northoff, in his work on the cortical midline self, has mapped the neural basis of this self-referentiality and shown that ego-processing becomes denser precisely at those nodes where the causal integration of the entire system also has its centre of gravity (Northoff et al. 2006, Northoff and Bermpohl 2004).

From the perspective developed here, the ego is the reflexive signature of the causal core, that is, the way in which the centre registers itself as a centre. It is not an additional instance alongside the self but its self-referential modulation. The difference between self and ego can be stated precisely. The self is internal asymmetry as such, that is, the fact of an integrated centre. The ego is internal asymmetry that images itself, that is, a centre that not only integrates but identifies this integration as its own. This explains why ego-processing is bound to more highly developed cortical structures, while the self can already be realised subcortically. Reflexive self-imaging requires the integration density of the default mode network, while basal integration is also possible in the brainstem.

The concept of a reflexive signature must not remain metaphorical at this point. Thermodynamically understood, it designates a second dissipative loop in which the integration dynamics of the core does not merely process environmental inputs but feeds its own integration state back into itself as input. Edelman's concept of reentrant signalling (Edelman 1989, Edelman and Tononi 2000) describes this recurrent feedback of neural groups that modulate themselves through their own activity. In the language of statistical physics, this formally corresponds to metastable dynamics as analysed by Tognoli and Kelso and by Deco and Jirsa (Tognoli and Kelso 2014, Deco and Jirsa 2012). Metastable heteroclinic channels are flow structures that hold the system in a recurrent cycle without collapsing into a static attractor. On this reading, the ego is the thermodynamic expression of such a recurrent dissipative loop, not an additional entity. It is the signature of a core that makes its own integration part of its integration. That this describes the structural precondition of the phenomenal self and not the phenomenal self itself has already been noted. The reconstruction remains deliberately structural at this point and forgoes an answer to the hard question concerning the qualia-character of self-reference.

6.3 Agency as outward action

Work by Patrick Haggard on the sense of agency shows that the experience of authorship of action is not bound to any one location but to the integration coherence between motor intention, efference copy, and sensory feedback (Haggard 2017, Moore and Obhi 2012). When this integration breaks down, for instance in schizophrenia or in lesions of the supplementary motor area, it is not movement that disappears but its attributability to the centre. The patient moves, but does not experience the movement as their own.

From the perspective of the causal core, agency is the outward-directed side of the same centre, that is, the way in which the centre reaches into the periphery and attributes this

reach to itself. It is neither a capacity alongside self and ego nor a retrospective attribution but a constitutive moment of the central structure. A centre that did not act outwards would be no centre. Agency is that aspect of the causal core which translates integration into executable directedness. This explains the empirical observation that disturbances of agency and disturbances of the self almost always appear together clinically. They are not two failures but two aspects of one failure.

6.4 Convergence in clinical findings

The belonging-together of the three aspects is empirically confirmed by the lesion network mapping studies of Michael Fox and colleagues (Fox 2018, Darby et al. 2018, Joutsa et al. 2022). Disturbances of self, ego, and agency arise from lesions at quite different anatomical sites, provided that these lesions affect the same functional node within the network. Two anatomically remote lesions can produce the same clinical syndrome when they functionally disrupt the same hub. This is the clinical demonstration that the three phenomena are not independent modules but aspects of a single integration structure. They break down together because they rest together on the same integration density.

The puzzle of the close coupling of self, ego, and agency is thereby dissolved. They are not three functions that mysteriously cooperate but three perspectives on the same integration centre. Their empirical correlation is not a surprise in need of explanation but a predictable consequence of their shared anchoring in the causal core.

7. Conclusion

The empirical situation supports a continuity thesis that does without a metaphysics of emergence and undercuts the classical dichotomies between biological and cognitive description. The same formal principle of flow-driven symmetry breaking with the formation of a central density zone manifests itself in prokaryotic polarity, in the Min oscillation, in biomolecular condensates, in neuronal directedness, in the laminar architecture of the cortex, in the rich club of the connectome, and in the broken detailed balance of brain dynamics. At several levels it has also taken on morphological form, first in the eukaryotic nucleus, later in the centralised nervous system, and finally in the vertebrate brain.

The five standard objections to such a continuity thesis can be met provided that the claim is formulated as a formal invariance under an evolutionarily conserved polarity machinery rather than as a metaphysical relation of identity. In this refined form, the model opens up the application to self, ego, and agency as three descriptive directions of the same causal core. The self is its internal structure, the ego its reflexive signature, agency its outward action. That these three phenomena are so tightly coupled empirically and collapse together under disturbances of central integration is not a puzzling coincidence but the predictable consequence of their common anchoring in the causal core.

The causal core thereby proves to be the thread of a biological continuity that reaches from the first membrane-enclosed autocatalytic cell to the self-referential integration dynamics of the human brain. It is not a special phenomenon of a particular organisational stage but the

constructional principle that has structurally shaped the evolution of the living from the outset, taking on a new morphological and functional form at every level of complexity.

References

Arimura, N. and Kaibuchi, K. (2007). Neuronal polarity: from extracellular signals to intracellular mechanisms. Nature Reviews Neuroscience 8, 194–205.

Baars, B. J. (1988). A Cognitive Theory of Consciousness. Cambridge: Cambridge University Press.

Banani, S. F., Lee, H. O., Hyman, A. A. and Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nature Reviews Molecular Cell Biology 18, 285–298.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P. and Friston, K. J. (2012). Canonical microcircuits for predictive coding. Neuron 76, 695–711.

Betti, L. and Hagmann, P. (2023). From connectome to cognition: architectural principles of the brain as a network. Current Opinion in Neurobiology 78, 102677.

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F. and Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science 324, 1729–1732.

Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N. and Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. Nature Reviews Neuroscience 17, 718–731.

Crossley, N. A., Mechelli, A., Scott, J., Carletti, F., Fox, P. T., McGuire, P. and Bullmore, E. T. (2014). The hubs of the human connectome are generally implicated in the anatomy of brain disorders. Brain 137, 2382–2395.

Damasio, A. (1999). The Feeling of What Happens. New York: Harcourt.

Damasio, A. (2010). Self Comes to Mind. New York: Pantheon.

Darby, R. R., Joutsa, J., Burke, M. J. and Fox, M. D. (2018). Lesion network localization of free will. Proceedings of the National Academy of Sciences 115, 10792–10797.

Deco, G., Cruzat, J., Cabral, J., Tagliazucchi, E., Laufs, H., Logothetis, N. K. and Kringelbach, M. L. (2019). Awakening: predicting external stimulation to force transitions between different brain states. Proceedings of the National Academy of Sciences 116, 18088–18097.

Deco, G. and Jirsa, V. K. (2012). Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. Journal of Neuroscience 32, 3366–3375.

Dehaene, S. and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. Neuron 70, 200–227.

Di Paolo, E., Buhrmann, T. and Barandiaran, X. (2017). Sensorimotor Life: An Enactive Proposal. Oxford: Oxford University Press.

Dotti, C. G., Sullivan, C. A. and Banker, G. A. (1988). The establishment of polarity by hippocampal neurons in culture. Journal of Neuroscience 8, 1454–1468.

Ebersbach, G. and Jacobs-Wagner, C. (2007). Exploration into the spatial and temporal mechanisms of bacterial polarity. Trends in Microbiology 15, 101–108.

Edelman, G. M. (1989). The Remembered Present: A Biological Theory of Consciousness. New York: Basic Books.

Edelman, G. M. and Tononi, G. (2000). A Universe of Consciousness: How Matter Becomes Imagination. New York: Basic Books.

Fox, M. D. (2018). Mapping symptoms to brain networks with the human connectome. New England Journal of Medicine 379, 2237–2245.

Frank, F. C. (1953). On spontaneous asymmetric synthesis. Biochimica et Biophysica Acta 11, 459–463.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11, 127–138.

Goehring, N. W. and Grill, S. W. (2013). Cell polarity: mechanochemical patterning. Trends in Cell Biology 23, 72–80.

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., Medaglia, J. D., Vettel, J. M., Miller, M. B., Grafton, S. T. and Bassett, D. S. (2015). Controllability of structural brain networks. Nature Communications 6, 8414.

Haggard, P. (2017). Sense of agency in the human brain. Nature Reviews Neuroscience 18, 196–207.

Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V. J., Meuli, R., Thiran, J. P. and Grant, P. E. (2010). White matter maturation reshapes structural connectivity in the late developing human brain. Proceedings of the National Academy of Sciences 107, 19067–19072.

Halatek, J. and Frey, E. (2018). Rethinking pattern formation in reaction-diffusion systems. Nature Physics 14, 507–514.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. Nature Reviews Neuroscience 6, 877–888.

Hyman, A. A., Weber, C. A. and Jülicher, F. (2014). Liquid-liquid phase separation in biology. Annual Review of Cell and Developmental Biology 30, 39–58.

Jacobs-Wagner, C. (2004). Regulatory proteins with a sense of direction: cell cycle signalling network in Caulobacter. Molecular Microbiology 51, 7–13.

Joutsa, J., Moussawi, K., Siddiqi, S. H., Abdolahi, A., Drew, W., Cohen, A. L., Ross, T. J., Haber, S. N., Stern, E. A. and Fox, M. D. (2022). Brain lesions disrupting addiction map to a common human brain circuit. Nature Medicine 28, 1249–1255.

Katz, L. C. and Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. Science 274, 1133–1138.

Kondepudi, D. K. and Asakura, K. (2001). Chiral autocatalysis, spontaneous symmetry breaking, and stochastic behavior. Accounts of Chemical Research 34, 946–954.

Kondepudi, D. K. and Prigogine, I. (1998). Modern Thermodynamics: From Heat Engines to Dissipative Structures. Chichester: Wiley.

Kringelbach, M. L. and Deco, G. (2020). Brain states and transitions: insights from computational neuroscience. Cell Reports 32, 108128.

Lane, N. (2015). The Vital Question: Why Is Life the Way It Is? London: Profile Books.

Lane, N. and Martin, W. (2010). The energetics of genome complexity. Nature 467, 929–934.

Lizier, J. T. (2013). The Local Information Dynamics of Distributed Computation in Complex Systems. Berlin: Springer.

Loose, M., Fischer-Friedrich, E., Ries, J., Kruse, K. and Schwille, P. (2008). Spatial regulators for bacterial cell division self-organize into surface waves in vitro. Science 320, 789–792.

Lynn, C. W., Cornblath, E. J., Papadopoulos, L., Bertolero, M. A. and Bassett, D. S. (2021). Broken detailed balance and entropy production in the human brain. Proceedings of the National Academy of Sciences 118, e2109889118.

Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. Behavioral and Brain Sciences 30, 63–134.

Moore, J. W. and Obhi, S. S. (2012). Intentional binding and the sense of agency: a review. Consciousness and Cognition 21, 546–561.

Northoff, G. and Bermpohl, F. (2004). Cortical midline structures and the self. Trends in Cognitive Sciences 8, 102–107.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H. and Panksepp, J. (2006). Self-referential processing in our brain. NeuroImage 31, 440–457.

Pasqualetti, F., Zampieri, S. and Bullo, F. (2014). Controllability metrics, limitations and algorithms for complex networks. IEEE Transactions on Control of Network Systems 1, 40–52.

Petersen, S. E. and Sporns, O. (2015). Brain networks and cognitive architectures. Neuron 88, 207–219.

Polleux, F. and Snider, W. (2010). Initiating and growing an axon. Cold Spring Harbor Perspectives in Biology 2, a001925.

Prigogine, I. and Nicolis, G. (1977). Self-Organization in Nonequilibrium Systems. New York: Wiley.

Raichle, M. E. (2015). The brain default mode network. Annual Review of Neuroscience 38, 433–447.

Schreiber, T. (2000). Measuring information transfer. Physical Review Letters 85, 461–464.

Seifert, U. (2012). Stochastic thermodynamics, fluctuation theorems and molecular machines. Reports on Progress in Physics 75, 126001.

Shapiro, L., McAdams, H. H. and Losick, R. (2009). Why and how bacteria localize proteins. Science 326, 1225–1228.

Shipp, S. (2007). Structure and function of the cerebral cortex. Current Biology 17, R443–R449.

Thompson, E. (2007). Mind in Life: Biology, Phenomenology, and the Sciences of Mind. Cambridge, MA: Harvard University Press.

Tognoli, E. and Kelso, J. A. S. (2014). The metastable brain. Neuron 81, 35–48.

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. Biological Bulletin 215, 216–242.

Tononi, G., Boly, M., Massimini, M. and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. Nature Reviews Neuroscience 17, 450–461.

van den Heuvel, M. P. and Sporns, O. (2011). Rich-club organization of the human connectome. Journal of Neuroscience 31, 15775–15786.

van den Heuvel, M. P., Kahn, R. S., Goñi, J. and Sporns, O. (2012). High-cost, high-capacity backbone for global brain communication. Proceedings of the National Academy of Sciences 109, 11372–11377.

Wedlich-Söldner, R. and Li, R. (2003). Spontaneous cell polarization: undermining determinism. Nature Cell Biology 5, 267–270.