The Placebo Effect as a Phenomenon of Autocatalytic Organization

Abstract

The standard neuroscientific account of the placebo effect identifies the neurochemical systems involved but fails to explain the constitutive transition from semantic context to physiological cascade. The dominant explanatory concept—expectation—does not resolve this gap; it merely relabels the explanandum. This paper proposes an alternative framework based on the theory of autocatalytic organization, drawing on the formal work of Rosen, Kauffman, and Montévil/Mossio. In an autocatalytically organized system, there is no translation between "meaning" and "physiology" because both are aspects of the same self-maintaining process. The placebo effect is reconceived as the integration of a perturbation into self-reinforcing cascades, governed by the topology of the autocatalytic network. This framework is applied to four empirical domains: psychoneuroimmunology (conditioned immunosuppression), open-label placebos, network neuroscience, and the placebome. Concrete, testable hypotheses are derived, including the prediction that perturbation-based measures of network non-factorizability should correlate with placebo magnitude. The clinical implications include a reframing of the drug–placebo distinction, a critique of the factorizability assumption underlying RCT design, and a rationale for individualized open-label placebo interventions.

The Explanatory Deficit of the Standard Account

Neuroscientific placebo research has made impressive empirical progress over the past two decades. It is now possible to identify precisely which neurochemical systems are involved in placebo responses: endogenous opioids, dopaminergic pathways in the ventral striatum, endocannabinoids, and cholecystokinin as an antagonist (Benedetti et al. 2005). The brain regions involved are well characterized, particularly the dorsolateral prefrontal cortex, the anterior cingulate cortex, and the periaqueductal gray (Wager et al. 2004). Conditioning and verbal suggestion activate different but overlapping mechanisms, and targeted pharmacological blockade (e.g., with naloxone) can selectively abolish specific placebo components (Amanzio and Benedetti 1999).

What the standard account fails to provide, however, is an explanation of the constitutive transition from what is typically described as a "semantic" impulse to a physiological cascade. The prevailing explanation holds that the patient *expects* improvement and that this expectation triggers neurochemical processes. But "expectation" here is not an explanatory concept; it is a reformulation of the explanandum. To say that expectation causes physiological changes is merely to affix a psychological label to the very transition that requires explanation. The question of how a meaning structure causally engages molecular pathways remains unanswered. Predictive processing (PP) shifts the problem by one step: it holds that the organism generates predictions and that placebo interventions modulate priors. But the question of how a semiotic content—trust in the physician, the significance of the white coat—*becomes* a prior remains conceptually unresolved. What operates here is an implicit dualism that treats semantic and physical levels as separate domains and must then postulate a mysterious bridging mechanism (a more differentiated engagement with PP follows below).

This deficit is not merely of philosophical interest. It has direct consequences for clinical research. If the standard explanation cannot specify *how* the transition from context to physiology works, it cannot systematically predict *when* and *in whom* placebo effects will occur. Clinical placebo research therefore operates largely descriptively: effects are observed, moderators catalogued, and biomarkers sought, without an explanatory model capable of integrating the heterogeneous findings.

The Autocatalytic Perspective: Perturbation Instead of Translation

A fundamentally different perspective emerges when the organism is consistently conceived as an autocatalytically organized system. The concept of autocatalysis has a precise meaning here that goes beyond its common metaphorical use as self-reinforcement. In Robert Rosen's theory of (M,R)-systems (Rosen 1991), it designates an organization in which every component of the system is produced by other components of the same system, including the catalysts that enable this production. The system is, in Rosen's formulation, "closed to efficient causation": it requires no external agent to maintain its organization. Stuart Kauffman formalized this idea in the theory of autocatalytic sets and showed that such systems arise spontaneously above a certain complexity threshold (Kauffman 1993). Montévil and Mossio (2015) developed the related but non-identical idea of "closure of

constraints," which describes how in biological systems the boundary conditions of processes are themselves products of processes that operate under those boundary conditions.

What these formal approaches share, and what is decisive for the placebo effect, is the following: in an autocatalytically closed system, there is no principled separation between "meaning" and "physiology," because both are aspects of the same self-maintaining process. What we describe as "meaning" is the relational structure through which a perturbation is taken up by the system and integrated into its self-maintenance dynamics. What we describe as "physiology" is the material realization of this dynamics at the molecular and cellular level. The distinction between the two is a distinction of descriptive perspective, not of the thing itself.

It is important to demarcate this position from two related but insufficient approaches. The first is the concept of allostasis (Sterling and Eyer 1988). Allostasis describes the adaptive adjustment of physiological setpoints in response to changing demands, thereby moving beyond the rigid homeostasis model. But allostasis remains within the regulatory paradigm: there is a system with defined variables and setpoints that can be shifted. In an autocatalytically organized system, by contrast, the variables themselves are products of the system. The "setpoints" are not adjusted; rather, the entire process architecture reorganizes. The placebo effect cannot be adequately described in the language of allostasis, because it does not consist in the shifting of setpoints but in the recruitment of cascades that concern the process architecture itself.

The second approach that requires a more differentiated engagement is predictive processing (PP). One could argue that PP already constitutes a formal implementation of autocatalytic principles, insofar as predictive signals might be understood as self-generated perturbations of the system. This reading, however, overlooks a decisive difference. PP works with a system that maintains a *model* of its environment and updates this model through prediction errors. This presupposes a distinction between the modeling system and the modeled world: the brain generates predictions *about* the world and corrects them on the basis of sensory evidence. In an autocatalytically organized system, this "about" does not exist. There is no "model" that can be distinguished from a "process"; the organization *is* the dynamics. PP displaces the dualism from semantics/physiology to model/world without resolving it. When PP says the placebo effect arises through the modulation of priors, it describes a regulatory operation within a modeling system. The question of how a semiotic content (trust in the

physician) *becomes* a prior remains unanswered, because PP presupposes the constitution of the model rather than explaining it. The autocatalytic framework, by contrast, does not ask how priors change but under what conditions of system organization a perturbation is integrated in a cascading manner—and it can answer this question without recourse to a model–world distinction.

The decisive theoretical step thus consists in abandoning the concept of "translation" between semantics and physiology. In an autocatalytically organized system, there is no translation because there are no two separate languages. There is a unitary process that appears as physicochemical or as meaningful depending on the descriptive perspective. The physician's coat does not "contain" meaning that must then be translated into neurochemistry. It enters as a perturbation into a system whose organization is already such that certain perturbation patterns trigger or reinforce certain autocatalytic loops.

A clarification is necessary here regarding the concept of perturbation itself. It would be a misunderstanding to treat the perturbation as an external input that impinges on a pre-existing system from outside. In an autocatalytically closed system, the boundary between system and environment is not given but is itself a product of the system's organization. The white coat is not an object "out there" that is then somehow "let in." It is a configuration of photons striking the retina, which becomes a perturbation only insofar as the system's organization is such that this configuration constitutes a difference that propagates through its self-maintaining loops. A system that lacks the relevant organizational history does not fail to "translate" the coat; the coat simply does not exist as a perturbation for that system. The perturbation, in other words, is not something that happens *to* the system but something that the system *constitutes as perturbation* by virtue of its own organization. This point, familiar from the theory of autopoiesis (Maturana and Varela 1980), has a specific consequence for the placebo case: it means that "topological fit" is not a match between two independent entities (perturbation and network) but a property of the system's organization alone.

This is not merely a linguistic correction. It changes the structure of the explanatory problem (Stegemann 2025a). The question is no longer: How does meaning become physiology? But rather: Under what conditions of system organization is a given perturbation integrated into self-reinforcing cascades? This reformulation has immediate consequences for clinical practice, as it shifts the focus from the patient's "expectation" (a variable that is clinically

difficult to manipulate) to the topology of the autocatalytic network (a variable that can in principle be addressed through the choice and design of the perturbation).

Conditions of Perturbation Integration

To answer the central question, one must specify the conditions under which an autocatalytically organized system processes perturbations in different ways.

Topological fit. Not every perturbation triggers self-reinforcing cascades. The white coat does not act universally but only in a system whose autocatalytic organization already provides points of connection for this specific perturbation. An organism that has never developed a link between the appearance of a physician and recovery processes will not be perturbed by the coat in the manner described as a placebo effect. This corresponds empirically to what conditioning research describes, but the crucial difference is that what is involved is not a learned "association" between two separate domains but a differentiation of the autocatalytic organization itself. The organism has not learned that a coat "means" healing; rather, its autocatalytic topology has differentiated such that certain sensory patterns potentiate certain self-maintenance cascades.

The clinical relevance of this point is considerable. Psychoneuroimmunology (PNI) has shown that peripheral immune functions can be modulated through classical conditioning. In a paradigmatic experimental design, Schedlowski, Engler, and colleagues repeatedly paired a gustatory stimulus with the immunosuppressant cyclosporine A; after the conditioning phase, the gustatory stimulus alone was sufficient to significantly reduce IL-2 production by T lymphocytes (Schedlowski and Enck 2015). The neuroanatomical pathways run through the insular cortex, the amygdala, and noradrenergic efferents to lymphatic tissue. Schedlowski is already testing this principle in clinical pilot projects with transplant patients, in which conditioned immunosuppression is intended to reduce the dosage of the immunosuppressant (Kirchhof et al. 2018). From the autocatalytic perspective, this is not a "placebo trick" but the targeted exploitation of topological fit: conditioning differentiates the autocatalytic topology such that the gustatory stimulus, as a perturbation, recruits the immunological loops.

A specific finding by Albring et al. (2012) is particularly instructive: a single re-exposure to the conditioned stimulus was insufficient to trigger the immunosuppressive response; only four re-exposures reliably produced the effect. This result can be interpreted as a threshold phenomenon. The perturbation must reach a certain intensity or repetition rate before the

autocatalytic organization converts it into a self-reinforcing cascade. Below this threshold, the perturbation is absorbed locally without recruiting the topologically more distant loops. Above the threshold, the dynamics tips into a cascading mode in which the perturbation traverses the entire nested network. For the clinical application of conditioned immunomodulation, this means that the dosing parameters (frequency and temporal distribution of re-exposures) are not merely quantitative variables but determine whether the system enters the cascading mode at all. The autocatalytic theory predicts a qualitative phase transition here, not a gradual increase, and this prediction is experimentally testable.

Non-factorizability. The more complex an autocatalytically organized system, the more nested its self-maintenance processes, and the less the individual loops can be isolated (Stegemann 2025a). In a highly nested system, a perturbation that apparently affects only a local loop can cascade to distant sub-processes because the loops are not factorizable from one another. The PNI finding illustrates this exemplarily: a gustatory stimulus modulates T-cell proliferation through cortical, subcortical, and autonomic systems that are in turn intertwined through feedback loops (e.g., cytokine-mediated signals from the immune system to the brain). This is precisely the kind of non-factorizable causality that the autocatalytic model predicts.

Network neuroscience confirms this picture at a different level of description. Wager and Atlas (2015) argued that the ventromedial prefrontal cortex (vmPFC) should not be understood as an isolated “placebo center” but as a hub of a network that generates a situational self-model. More recent work on functional connectivity shows that the Default Mode Network (DMN) in particular plays a central role in generating placebo effects. Its coupling with the pain network, the reward system, and the autonomic nervous system exhibits precisely the kind of nested organization that the autocatalytic model describes as constitutive. Simultaneously, it has been observed that during placebo effects, the Ventral Attention Network (VAN) is deactivated (Wager et al. 2004), suggesting that the perturbation does not merely recruit local loops but reconfigures the global processing architecture of the system.

The decisive point is this: network neuroscience already describes the non-factorizability of the placebo response empirically but lacks a theoretical framework that identifies it as a constitutive feature of the phenomenon. Researchers observe that placebo effects recruit multiple, non-isolable networks, and they interpret this as a methodological problem: How do

we isolate the “real” mechanism? The autocatalytic model reverses this perspective: non-factorizability is not an obstacle to explanation but the explanation itself. Placebo effects occur *because* the causal loops of the system are not isolable from one another. The search for the “real” placebo mechanism behind the network activity is as misguided as the search for the “real” translation mechanism between meaning and physiology.

From this follows a concrete, empirically testable hypothesis: the degree of functional connectivity between the networks involved in the placebo response (as a proxy for the degree of non-factorizability) should correlate with the magnitude of the placebo effect. Individuals with higher network integration should exhibit stronger and broader placebo effects than individuals whose networks are more segregated.

A methodological clarification is necessary here, because resting-state fMRI alone provides only an image of static connectivity and thus does not capture what the theory actually requires: the dynamic response of the system to perturbations. Autocatalytic topology is not a structure that can be fully mapped at rest but a dispositional property that manifests only in the reaction to specific perturbations. A perturbation-based approach would therefore be methodologically more adequate. The combination of transcranial magnetic stimulation and functional imaging (TMS-fMRI) allows targeted perturbation of network hubs and real-time observation of activation spread across the network (Bergmann et al. 2021). The pattern of this spread—in particular, whether a local perturbation remains local or triggers global cascades—would be a direct empirical indicator of the degree of non-factorizability. Pharmacological challenges (e.g., low-dose naloxone as a perturbation of the endogenous opioid system) could be used analogously to test the cascading propensity of specific network loops. The crucial point is that the theory not only reinterprets existing methods but motivates a concrete methodological program: measuring non-factorizability requires perturbation-based network analysis, not mere correlation measures.

Reflexivity. Human organisms are characterized by the fact that their autocatalytic organization includes a reflexive level: the processes of the system can be related to themselves. A human being does not merely take a pill; they know that they are taking a pill, and this knowledge is itself a perturbation that enters the autocatalytic organization. Reflexivity potentiates the effect because it opens an additional feedback loop: the expectation of improvement generates subtle physiological changes, which in turn enter the system as perturbations and reinforce the expectation. At the same time, reflexivity explains

the nocebo effect as a structural mirror image: a downward spiral that destabilizes existing self-maintenance cascades rather than reinforcing them.

Open-Label Placebos: Test Case and Clinical Perspective

A particularly instructive test case for the autocatalytic perspective is provided by open-label placebos (OLP). Kaptchuk et al. (2010) showed that patients with irritable bowel syndrome experienced significant improvement despite being explicitly informed that they were receiving a placebo with no active ingredient. Since then, the finding has been replicated for chronic low back pain, cancer-related fatigue, allergic rhinitis, depression, test anxiety, and migraine (von Wernsdorff et al. 2021; Schaefer et al. 2016, 2019, 2024). Meta-analyses confirm moderate to large effect sizes compared to no-treatment groups (Spille et al. 2023).

For the standard explanation via "expectation," OLP pose a massive conceptual problem. How can someone "expect" something they know to be inert? The usual explanatory attempts resort to auxiliary constructions: "unconscious conditioning," "implicit expectation" generated by the ritual, or "expectation modulation" through the accompanying explanation. These explanations are not wrong, but their very plurality and heterogeneity show that the concept of expectation is no longer viable as a unified explanans. It is stretched ad hoc to cover findings it cannot explain in its original meaning.

From the autocatalytic perspective, the problem dissolves because it does not arise as a problem in the first place. The OLP does not work because the patient "expects" it to work but because the ritual of ingestion enters as a perturbation into an autocatalytic network whose topology has differentiated over a lifetime of experience with medical contexts. The conscious knowledge that the pill is inert is itself only one among many perturbations simultaneously acting on the system. It competes with deeply inscribed autocatalytic loops that have developed over decades of experience with physicians, pills, and recovery episodes. Non-factorizability explains why reflexive knowledge cannot "switch off" the sub-cognitive loops: in a non-factorizable system, there is no isolated causal pathway from the reflexive representation to the sub-cognitive loops that would allow the former to selectively inhibit the latter. The system processes the perturbation "it is a placebo" and the perturbation "someone in a white coat is giving me something to take" not independently of each other but as an entangled whole whose outcome is determined by the overall topology. In a factorizable system, by contrast, the knowledge that a substance is inert could in principle deactivate the

response; that OLP work is itself evidence for the non-factorizability of the underlying causal structure.

This interpretation is empirically supported by a remarkable finding. Schaefer et al. (2021) found that the actual placebo effect in OLP did not correlate with participants' conscious expectations but rather with a dispositional variable they termed "belief in placebos." The authors conclude that OLP operate predominantly through unconscious processes. A more recent study (Schaefer, Leupold, and Enge 2025) replicates this pattern and shows that neither the route of administration (oral vs. topical) nor the explicit expectation measure predicts the OLP effect, but the dispositional conviction does. In the language of the autocatalytic model: it is the topology of the network as differentiated over the life history (the "disposition"), not the current reflexive representation (the "expectation"), that determines whether a perturbation triggers self-reinforcing cascades.

The clinical implication is far-reaching: OLP could be deployed as an ethically unproblematic therapeutic instrument, precisely *because* they require no deception mechanism. The autocatalytic theory explains why this works and simultaneously provides criteria for optimization: efficacy should depend not on the persuasiveness of the verbal explanation (as expectation theory would suggest) but on the topological fit of the ritual to the individual patient's autocatalytic organization. This means that OLP should be understood not as a uniform intervention but as a perturbation to be individualized topologically.

The Placebome: Genetic Architecture of Network Topology

Hall, Loscalzo, and Kaptchuk (2015) have shown that genetic variants correlate with placebo responsivity, particularly polymorphisms in the COMT gene (catechol-O-methyltransferase, which regulates the degradation of dopamine and norepinephrine), the OPRM1 gene (μ-opioid receptor), and genes involved in serotonergic and endocannabinoid systems. Patients with the COMT met/met genotype, which is associated with slower dopamine degradation and thus higher basal dopamine levels, consistently show stronger placebo effects (Hall et al. 2012). Wang et al. (2017) complemented this with a network analysis of the genomic basis of the placebo effect, showing that the genes involved do not form an isolated group but a functionally interconnected network.

Within the standard account, this finding is typically interpreted as meaning that genetic variation determines the individual's "neurochemical endowment" and thus their "capacity"

to respond to expectations. This interpretation presupposes the problematic expectation concept and treats the genetic substrate as a mere enabling condition for a psychological mechanism.

The autocatalytic perspective offers a different reading. Genetic variation does not determine the "capacity for expectation" but shapes the topology of the autocatalytic network itself. A polymorphism in the COMT gene alters the speed of dopamine degradation and thereby the dynamic properties of dopaminergic loops, which are in turn embedded in neuroendocrine, autonomic, and immunological cascades. Genetic variation thus modulates not a psychological mechanism but the architecture of the autocatalytic network and thereby the "connectability" of certain perturbations.

This interpretation has a concrete advantage: it explains why genetic placebo responsivity appears consistently across different placebo paradigms (Hall et al. 2015). If the placebome modulated a specific "expectation mechanism," one would expect it to operate only with certain types of placebos. If, however, it shapes the topology of the autocatalytic network, then it influences the general cascading propensity of the system. The practical consequence: placebome genotyping could serve as an instrument for predicting individual responsivity—not because it identifies a "placebo responder" but because it provides information about the cascading propensity of the autocatalytic network. In combination with resting-state fMRI data (as a proxy for functional network topology) and individual learning history, this could enable a multidimensional prediction approach.

Non-Teleological Clarification

A conceptual clarification is necessary at this point. An autocatalytically organized organism is not a teleological system. It does not "pursue" a goal and is not "directed" toward self-maintenance in the sense that self-maintenance would be a purpose. Rather, self-maintenance is the condition of its existence as an organized system: as long as the autocatalytic processes run, the system exists; when they cease, the organization dissolves. The organism does not "use" the white coat "for self-healing," as though it were pursuing a strategy. Rather, the organization of the system is such that certain perturbations enter self-reinforcing cascades that we retrospectively describe as "healing." The teleological interpretation generates an explanans (the will or goal of healing) that would itself require explanation and leads to a

regress. The autocatalytic interpretation explains the same finding solely from the organizational logic of the system.

The Drug–Placebo Distinction as a Descriptive Artifact

Perhaps the most far-reaching consequence of the autocatalytic perspective concerns the relationship between placebo and pharmacological action. Benedetti and colleagues (2005) have shown that placebos and active substances partially activate the same neurochemical pathways and recruit the same receptor systems. In the standard account, this appears as a curiosity: How can a "sham medication" activate the same receptors as a "real" medication? In the autocatalytic model, the finding is to be expected: both the pharmacological substance and the contextual factors are perturbations of the same autocatalytic network. They differ in their entry point (the substance engages receptors directly; the context perturbs through sensory and cognitive loops), but because the network is non-factorizable, both perturbations converge on similar cascades.

Mayberg et al. (2002) showed by means of FDG-PET that fluoxetine and placebo produced a shared pattern of cortical increases and limbic-paralimbic decreases in depression patients, with fluoxetine producing additional changes in brainstem, hippocampus, insula, and caudate. This is to be expected within the autocatalytic model: both perturbations engage the same network, but the pharmacological perturbation has a broader entry point due to its direct receptor binding.

The strict distinction between "pharmacological" and "placebo" effects thus proves to be an artifact of a description that decomposes the system into independent components it does not possess. The consequences for clinical research are considerable. The design of the randomized controlled trial (RCT) rests on the assumption that one can isolate the "pure" drug effect by subtracting the "placebo component." If, however, drug and placebo are perturbations of the same non-factorizable network, then the subtraction is not a clean operation because the interaction between the two perturbations is non-linear. This does not mean that RCTs are worthless, but it does mean that their logic rests on a factorizability assumption that the autocatalytic model identifies as empirically inaccurate. Rather than asking "Does the drug work?", one would need to ask: Which combination of perturbations (pharmacological and contextual) produces the strongest self-reinforcing cascade for a given autocatalytic topology?

Consciousness as a Special Case: Connection to Causal Collapse Theory

The autocatalytic analysis of the placebo effect points beyond its immediate subject matter. If consciousness, as proposed within the framework of causal collapse theory (Stegemann 2025b), is understood as the epistemic interior of a non-factorizable causal system, then the placebo effect would be a special case of a more general phenomenon: that conscious systems, by virtue of their non-factorizability, exhibit global responses to local perturbations that do not occur in factorizable systems. This could explain why the placebo effect increases in complexity and strength to the extent that the system's organization includes reflexive loops, and why purely pharmacological models that treat the organism as an aggregate of mutually independent subsystems systematically underestimate the effect.

Scope and Limitations

The framework proposed here is a theoretical reorientation, not a mechanistic model. It identifies the conditions under which placebo effects occur (topological fit, non-factorizability, reflexivity) and derives testable hypotheses, but it does not specify the molecular or synaptic processes through which these conditions are realized. This is a deliberate limitation, not an oversight. Rosen's (M,R)-systems do not specify which molecules perform which catalysis; Kauffman's autocatalytic sets do not enumerate which peptides participate. The formal theory operates at the level of organizational principles, and the task of identifying the material realization of these principles in specific placebo paradigms falls to empirical research. The present paper reorients that research by specifying what it should look for (cascading propensity upon perturbation, threshold phenomena, topological individuality) rather than what it has been looking for (expectation mechanisms, isolated neural correlates, single mediating pathways).

Two further limitations should be noted. First, the concept of "topological fit" is itself in need of empirical specification. The paper establishes that it is the system's organization, not a match between two independent entities, that determines whether a perturbation cascades. But the question of how specific organizational histories (conditioning, developmental experience, cultural embedding) inscribe themselves into the topology of the autocatalytic network remains open and constitutes a natural next step for a research program that bridges the present framework with the molecular neuroscience of learning and plasticity. Second, the claim that non-factorizability explains why certain perturbations cascade while others are

locally absorbed requires further differentiation. Not all non-factorizability is alike; future work will need to distinguish types of non-factorizability (hierarchically nested vs. reciprocally coupled, strongly vs. weakly integrated) and to develop perturbation-based empirical measures that can discriminate among them.

Summary and Outlook

The perspective developed here effects a shift in explanatory level. Rather than asking how meaning is translated into physiology, it asks under what conditions of system organization perturbations are integrated into self-reinforcing or self-destabilizing cascades. The placebo effect appears not as a puzzling borderline case between mind and body but as a regular phenomenon of autocatalytic organization.

The empirical findings converge: psychoneuroimmunology shows that the supposedly separate subsystems are not isolable (Schedlowski and Enck 2015). Open-label placebos show that the concept of expectation does not hold as a bridging mechanism (Kaptchuk et al. 2010; Schaefer et al. 2021). Network neuroscience shows that the empirical reality consists of non-factorizable networks (Wager and Atlas 2015). The placebome shows that genetic variation shapes the network topology itself (Hall et al. 2015). All of these findings are not merely explicable within the autocatalytic model but are to be expected.

The practical relevance of this theoretical move lies in three points. First, the therapeutic context (ritual, setting, physician–patient relationship) is not a confound to be controlled for but part of the efficacy itself. Clinical practice should optimize the interaction between pharmacological and contextual perturbations rather than attempting to isolate one from the other. Second, the prediction of individual placebo responsivity becomes possible when the topology of the autocatalytic network is assessed multidimensionally: placebome genotyping, perturbation-based network analysis (TMS-fMRI, pharmacological challenges), and individual learning history (conditioning paradigms). Third, open-label placebos and conditioned immunomodulation can be deployed as rational therapeutic instruments whose parameters (dosage, timing, context design) are derivable from autocatalytic theory and can be optimized in a theory-guided manner.

The puzzling character of the placebo effect dissolves not by finding the mechanism through which meaning becomes causal but by recognizing the question itself as an artifact of a dualistic framing that becomes moot within the autocatalytic paradigm.

References

Albring, A., Wendt, L., Benson, S., Witzke, O., Kribben, A., Engler, H., & Schedlowski, M. (2012). Placebo effects on the immune response in humans: The role of learning and expectation. PLoS ONE, 7(11), e49477.

Amanzio, M., & Benedetti, F. (1999). Neuropharmacological dissection of placebo analgesia: Expectation-activated opioid systems versus conditioning-activated specific subsystems. Journal of Neuroscience, 19(1), 484–494.

Benedetti, F., Mayberg, H. S., Wager, T. D., Stohler, C. S., & Zubieta, J. K. (2005). Neurobiological mechanisms of the placebo effect. Journal of Neuroscience, 25(45), 10390–10402.

Bergmann, T. O., Varatheeswaran, R., Hanlon, C. A., Madsen, K. H., Thielscher, A., & Siebner, H. R. (2021). Concurrent TMS-fMRI for causal network perturbation and proof of target engagement. NeuroImage, 237, 118093.

Hall, K. T., Lembo, A. J., Kirsch, I., Ziogas, D. C., Douaiher, J., Jensen, K. B., Conboy, L. A., Kelley, J. M., Kokkotou, E., & Kaptchuk, T. J. (2012). Catechol-O-methyltransferase val158met polymorphism predicts placebo effect in irritable bowel syndrome. PLoS ONE, 7(10), e48135.

Hall, K. T., Loscalzo, J., & Kaptchuk, T. J. (2015). Genetics and the placebo effect: The placebome. Trends in Molecular Medicine, 21(5), 285–294.

Kaptchuk, T. J., Friedlander, E., Kelley, J. M., Sanchez, M. N., Kokkotou, E., Singer, J. P., Kowalczykowski, M., Miller, F. G., Kirsch, I., & Lembo, A. J. (2010). Placebos without deception: A randomized controlled trial in irritable bowel syndrome. PLoS ONE, 5(12), e15591.

Kauffman, S. A. (1993). The origins of order: Self-organization and selection in evolution. Oxford University Press.

Kirchhof, J., Petrakova, L., Brinkhoff, A., Benson, S., Schmidt, J., Engler, H., Witzke, O., & Schedlowski, M. (2018). Learned immunosuppressive placebo responses in renal transplant patients. Proceedings of the National Academy of Sciences, 115(16), 4223–4227.

Maturana, H. R., & Varela, F. J. (1980). Autopoiesis and cognition: The realization of the living. Reidel.

Mayberg, H. S., Silva, J. A., Brannan, S. K., Tekell, J. L., Mahurin, R. K., McGinnis, S., & Jerabek, P. A. (2002). The functional neuroanatomy of the placebo effect. American Journal of Psychiatry, 159(5), 728–737.

Montévil, M., & Mossio, M. (2015). Biological organisation as closure of constraints. Journal of Theoretical Biology, 372, 179–191.

Rosen, R. (1991). Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life. Columbia University Press.

Schaefer, M., Harke, R., & Denke, C. (2016). Open-label placebos improve symptoms in allergic rhinitis. Psychotherapy and Psychosomatics, 85(6), 373–374.

Schaefer, M., Denke, C., Harke, R., Olk, N., Erkovan, M., & Enge, S. (2019). Open-label placebos reduce test anxiety and improve self-management skills: A randomized-controlled trial. Scientific Reports, 9, 13317.

Schaefer, M., Hellmann-Regen, J., & Enge, S. (2021). Effects of open-label placebos on state anxiety and glucocorticoid stress responses. Brain Sciences, 11(4), 508.

Schaefer, M., & Enge, S. (2024). Open-label placebos enhance test performance and reduce anxiety in learner drivers: A randomized controlled trial. Scientific Reports, 14, 6684.

Schaefer, M., Leupold, C., & Enge, S. (2025). Roles of administration route, expectation, and belief in placebos in a randomized controlled trial with open-label placebos. Scientific Reports, 15, 41197.

Schedlowski, M., & Enck, P. (2015). Neuro-bio-behavioral mechanisms of placebo and nocebo responses: Implications for clinical trials and clinical practice. Pharmacological Reviews, 67(3), 697–730.

Spille, L., Fendel, J. C., Seuling, P. D., Göritz, A. S., & Schmidt, S. (2023). Open-label placebos: A systematic review and meta-analysis of experimental studies with non-clinical samples. Scientific Reports, 13, 3640.

Stegemann, W. (2025a). The mechanism of phenomenal experience. Zenodo. https://doi.org/10.5281/zenodo.17698505

Stegemann, W. (2025b). Integrated model of causal collapse. Zenodo. https://doi.org/10.5281/zenodo.15350050

Sterling, P., & Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. In S. Fisher & J. Reason (Eds.), Handbook of life stress, cognition and health (pp. 629–649). Wiley.

von Wernsdorff, M., Loef, M., Tuschen-Caffier, B., & Schmidt, S. (2021). Effects of open-label placebos in clinical trials: A systematic review and meta-analysis. Scientific Reports, 11, 3855.

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., Kosslyn, S. M., Rose, R. M., & Cohen, J. D. (2004). Placebo-induced changes in fMRI in the anticipation and experience of pain. Science, 303(5661), 1162–1167.

Wager, T. D., & Atlas, L. Y. (2015). The neuroscience of placebo effects: Connecting context, learning and health. Nature Reviews Neuroscience, 16(7), 403–418.

Wang, R. S., Hall, K. T., Giulianini, F., Passow, D., Kaptchuk, T. J., & Loscalzo, J. (2017). Network analysis of the genomic basis of the placebo effect. JCI Insight, 2(11), e93911.