

Paper 473v3 Constructing Lived Cognition

[author names and references removed by editor to protect the integrity of the reviewing process]

Structured Abstract

Paper type: Synthetic.

Background(s): Cognitive science.

Approach: Enaction.

Context: Enactivism seeks to naturalise cognition, development and evolution as exemplars of a single meaning-making process, enaction, in autonomous systems.

Problem: Ezequiel Di Paolo proposes adaptivity as a sufficient naturalising condition for enactive autonomy, yet this condition fails to naturalise the normative distinction between adaptive and disruptive variation. The naturalisability of autonomy therefore remains an open issue.

Method: Substituting adaptivity by an agency criterion shifts explanatory focus to autonomous systems' causal irreducibility, entailing dynamical coordination of their structural relations. Coordination induces a naturalised basis for meaning-making in terms of narrative stabilisation: the stabilisation of coordinated narratives from stochastically varying structure. Narrative stabilisation is a naturalised selection process that regulates, not frequencies of replicating structures, but the stability of mutually coordinated narratives, which thereby attain normativity and intrinsic intentionality.

Results: A computer simulation demonstrates the ability of narrative stabilisation to implement normative adaptation of narratives, thereby fulfilling Di Paolo's criteria for enactive autonomy.

Implications: The emergence of autonomous societies through narrative stabilisation offers a naturalised, Kantian perspective on our scientific and ethical relationship to society and environment.

Constructivist content: The paper proposes a naturalised account of enaction.

Key Words: Adaptivity, agency, autonomy, cognition, enaction, learning, narrative stabilisation, naturalisation, normativity, semiosis.

Introduction: Autonomy characterises life

1. If, as organisms, we wish to understand and nurture our relationship to ourselves and to our physical and biological environment, we surely need to naturalise the notions of life, cognition and ethical relationship. That is, we need to explain, with no causal gaps, how normative, teleological living systems arise from and relate to the non-normative, non-teleological structures of purely physical and chemical relations accepted by the community of science. On the one hand, a naturalised account of life should help us to classify living systems from prokaryotes through symbioses and ecosystems to potential xeno-organisms. On the other, naturalisability constitutes a requirement on the philosophy of biology that may help us to diagnose and nurture the structural relations and functional organisation of ailing or healthy biological systems.

2. In this paper, I propose a naturalisation of life following this line of argument:

- A. Characterise what we mean by Life;
- B. Analyse the constraints that this characterisation imposes upon any putatively natural implementation; and
- C. Demonstrate some natural implementation that satisfies these constraints, and in addition satisfies the above characterisation of life.

For example, toward the end of the eighteenth century, Immanuel Kant (1996b) characterised (A) life in terms of *autonomy* – the choice to act independently of external causes. Humberto Maturana and Francisco Varela (1987) analysed (B) autonomy into the constraint that living systems stably maintain their own functional organisation. Ezequiel Di Paolo (2018) demonstrated (C) that adaptive systems satisfy this constraint by producing and maintaining their functional organisation against the predatory influence of perturbative variations in their bodily structure.

3. Yet this explanatory thread contains significant causal gaps. What do we mean, when we say that organisms *choose*, rather than algorithmically compute, their actions? How might *functional organisation* have first constituted itself out of non-functional physical and chemical structures? Based on which natural, normative criteria do living systems come to choose *adaptive* over potentially catastrophic variations of their structure?

4. I seek in this article to close these explanatory gaps and so approach more closely a seamlessly naturalised account of living and cognising. To be more precise, I shall argue the above steps A to C in this sequence:

- autonomy distinguishes choosing from computing;
- choice irreducibly constructs meaning;
- meaning is intentioned narrative;
- intention selects narratives;
- stabilised narratives are intrinsically intentioned; and
- stabilised narratives are autonomous.

Autonomy distinguishes choosing from computing

5. Kant's Critical Philosophy sought to account for humans' ability to combine two distinct modes of cognition: the *theoretical* logic of structural categories, evaluated with respect to proof and evidence; and the *practicalities* of dynamically choosing, evaluated with respect to normative judgement. For example, I may theoretically understand that meat production undermines both animal and environmental welfare; yet in order to choose in practice what to eat today, I must embrace or reject animal and environmental welfare, on the basis of some norm.

6. Kant (1996a) synthesised these two modes under the term *autonomy*. While Kant never speaks of autonomous systems, he defines a choice or action as autonomous if it is not dictated by foreign (i.e., non-self) authority. For example, my autonomous choice of food may be informed by a vegetarianism principle that is dictated by neither my constitutive structure nor my cultural context. To decouple choice from authority, Kant requires it to be *universalisable* across some collection of individuals in the following sense. As a human individual, I am structurally able to eat a vegetarian diet, since I share in the digestive structure of all humans; and I am free to choose a vegetarian diet, since, unlike a carnivorous diet, the dynamical consequences of this vegetarian choice will not undermine the ecological ground of my own existence, even if all humans made the same choice.

7. This definition seems at first strangely disembodied, yet it in fact sums up Kant's entire understanding of organismic embodiment. For while an isolated computing element is capable of *deciding*, only a tribunal of peers can *choose*, or pronounce judgement. Kant's (1996c, §40, 5:294) universalisability criterion requires that choosing fulfils three necessary preconditions: each individual acts according to its own principles; any individual is capable of adopting those same principles; and the integrated influence of all individuals' actions maintains the collective's existence. Autonomous choice is the *sensus communis* of a unitary collective of individuals.

8. Take, for example, a termite colony. Each worker termite responds to its local pheromone environment according to its own rules for manipulating wood mulch and pheromone; any termite possesses the physiological wherewithal to act according those same rules; and if all termites in the colony implement these deposition rules, their integrated consequences construct a non-local field of solidified mulch that constitutes an efficiently ventilated termite mound. This termite-friendly, constructed niche is the choice, the non-dictated judgement, of the entire termite collective.

9. Kant thus acknowledges the need for a dual-level account of autonomous action that respects both unitary organisation and structural components. The choosing system's organisation (the mound) constrains the operations of its constituent structures (termites) by choosing principles for action that do not undermine this organisation. If sufficient numbers of termites went rogue and operated according to incongruent

principles, this would undermine the integrity of the mound niche, which therefore possesses the agency to downwardly select the principles of its constituent termites on their ability to maintain this integrity.

10. Kant's dual-level requirement constrains quite tightly my naturalisation argument in this paper, which I will now sketch briefly. To evade the charge of teleology, I need to explain how this downward selection naturalises the *function* of a principle of action. Both Maturana and Varela's (1987) notion of autopoiesis, and Varela, Evan Thompson and Eleanor Rosch's (1991) later notion of enaction, define function in terms of integrity: principles fulfil the function of maintaining the autopoietic or enactive system's organisational integrity. Now, Brian Goodwin (1994) reinterpreted evolutionary selection as dynamical *stability* of organisation and the genetic structures that establish it; while Elliot Sober and David Sloan Wilson (1998) defined dual-level selection in terms of dependence upon a collective commons. Taken together, these ideas suggest the proposal (to be elaborated in this paper) that *the structural relations of an autonomous system collectively establish a unitary organisation, which itself stabilises those structural relations on their ability to maintain that self-same organisation.*

11. Further, to explain the emergence of autonomously acting systems, this dual-level account must describe the scaffolding (Hoffmeyer 2010) of structure-organisation principles that enact normatively useful, contextualised ways of behaving. Now, contributions by Geoffrey Hinton and Steven Nowlan (1987) and by Mary Jane West-Eberhard (2003) strongly indicate that searching for fixed, structural solutions to a problem is prohibitively less efficient than searching for processes that generate such solutions *exploratively*. As an illustration, when an expert guitarist breaks a fingernail, she at first mis-picks the strings, yet the quality of her playing quickly recovers. This is because her expertise consists not in fixed, off-the-peg fingering solutions, but rather in her ability to explore afresh the affordances of each new guitar-playing experience (Smith & Thelen 1993), while also plastically adapting the principles that she brings to this experience. I shall argue that such stabilised principles correspond to what Jerome Bruner (1990) calls *narratives*: intentioned explorations of how to resolve unexpected circumstances.

12. Bruner proposed narratives as fundamental, skill-based components of cognition, and the stabilisation of such narratives during learning will be our key to naturalising normativity. I will argue that the stabilisation of a narrative recursively entails its own intrinsic intentionality, making it a principle for fully naturalised autonomous action. Stabilisation is therefore the single norm of action underlying Kant's categorical imperative: *My action is adaptive if it stabilises the niche on which my capability to perform that action depends, taking into account the actions of all agents and components encompassed by that niche.*

13. This recursion clarifies the deep relationship between cognition and life. Cognition is not a tool that living systems employ, but is rather the natural, co-dependent arising that we call life. Biological organisms narrate their own autonomous identity by

skilfully surviving in the face of precarious experience. This skilful surviving is what we call cognition: organisms attend to experiential events; interpret those events using prior structural relations; develop personally relevant narrative meanings for these relations; abstract new structural relations that generalise across these meanings; and apply these relations through action. This cycle of using-abstracting relations through narrative stabilisation constitutes living systems' defining ability to enact (Varela, Thompson & Rosch 1991) a biological identity for whom events have meaning.

14. Throughout the article, I use the term *structure* to describe sets of relations such as relative location, orientation or time that are separable in the sense that we can imagine varying one relation without affecting other relations in the structure. Separability requires that structural variation is local: we can move or replace one chair, cell or allele independently of others. Locality implies that structural dynamics are digital, for in order to change a structure over time, we must perform a countable sequence of operations that each changes a particular relation within that structure. Operations may possess internal continuity, as when we carry one chair continuously to a new position in the room, but this internal continuity is independent of the next chair we move, partitioning our furniture rearranging into a sequence of discrete (i.e., local and digital) operations. Such a sequence of discrete structural operations is a computation, and computation is determined not by continuous dynamical time, but by digital sequence.

15. While structures pervade our everyday experience, the importance in physics of non-local minimisation principles suggests very strongly that something beyond purely local, discrete, structural relations underlies that experience. Stuart Kelso and David Engström (2006) use the term *coordination* to describe the recursive dynamics by which structures condition their own unitary organisation. For example, coordination binds protein folding, entangled quantum states or the attractor dynamics of guitar strings, hearts and hurricanes into organised unities. Coordination is characterised by continuity and non-locality, and often mediated by fields. A folding protein, an entangled quantum state or a brewing hurricane develops over continuous time and defies analysis into computational operations. Similarly, the trajectory of a single peptide in a folding protein, measurements on an entangled particle state, or the contraction of a single myocardial cell, are all non-local in the sense that they are coordinated within an overarching system organisation.

16. This usage differs slightly, but significantly, from that of Maturana and Varela (1987), for whom structure establishes (rather than conditions) a system's organisation. The important distinction is that we require both structure *and* coordination to establish organisation. For example, autopoietic organisation requires both molecules *and* their electromagnetic coordination, and this coordination is precisely what enables us to distinguish the dual levels of structure and organisation that will characterise narrative, enactive dynamics.

Choice irreducibly constructs meaning

17. A convenient entry-point for the naturalisation argument is Di Paolo's (2018) assertion that enactivism is the naturalistic, *non-reductive* study of living systems as essentially *embodied*. I take ...

- *non-reductive* to mean we accept that it may be impossible to explain cognitive, emotional and social phenomena in terms of “nothing but” their natural constitutive structure.
- *embodied* to mean that the nature of life and mind depends critically upon the kind of structure in which they are implemented. This contrasts with the functionalist view that mind consists in computational operations implemented on arbitrary hardware structures.

18. Enactivism's non-reductive naturalism draws upon the assumption that an unbroken explanatory path links living, mental, and social phenomena. This “life–mind continuity thesis” replaces the independent descriptive levels of chemistry, biology, psychology, sociology by the single assertion that living bodies are characterised at all levels by *adaptive autonomy*, which Di Paolo defines as requiring three properties of a dynamical system:

- *Operational closure*: The system's internal structural relations and operations collectively produce a stable, closed identity. Closure means that each relation of the system codetermines, and is codetermined by, some operations in the system; and each operation codetermines, and is codetermined by, some relations in the system.
- *Structural coupling*: This closed identity stands in reciprocal dynamical relationship with its environment, and must therefore distinguish itself from that environment by playing an indispensable role in codetermining its own reaction to any perturbations arriving from the environment. This ensures that perturbations have only a permissive, not a prescriptive, effect on the system's dynamics: the system's reaction to these perturbations is not passive, but rather an active response.
- *Adaptivity*: Both system and environment change reciprocally under the perturbative influence of each on the other. Due to the precarious, far-from-equilibrium nature of environmental relations, an autonomous system needs to adapt – to make decisions about which environmental matter and energy flows promote its self-production or endanger its self-distinction, and act accordingly.

19. Di Paolo focusses first on the two conditions of operational closure and structural coupling, characterising them respectively as conditions of self-production and self-distinction. These conditions drew originally on Maturana and Varela's (1980) biochemically motivated definition of *autopoiesis*; taken together, production and distinction naturalise the ability of an autopoietic system to choose unitarily, in terms of preserving its structure against environmental perturbation. While we might at first assume autonomy to be simply the generalisation of autopoiesis to non-biochemical

dynamical systems, the instantaneity of autopoiesis makes it unable to specify how a system historically becomes capable of autopoiesis. For this reason, Di Paolo introduces the third, historical, condition of adaptivity to pin down an idea expressed earlier by Varela (2000: 447f):

“[T]he notion of perturbation during structural coupling does not adequately take into account the regularities that emerge from a history of interaction [with the environment ...]. In these last few years I have developed an explicit alternative that [makes ...] historical reciprocity the key of a co-definition between an ‘autonomous’ system and its environment. It is what I propose to call the perspective of enaction in biology and cognitive science.” (Translated in Di Paolo 2018: 83).

20. Given the precarious nature of everyday relations with its environment, an autonomous system can never afford to relax in stable states, equilibria or even stationary probability distributions of states, but must *enact* the tools of its own survival out of its ongoing life-situation. That is, it monitors its boundaries and current state, continually adapting its state to maintain its own viability.

21. This adaptive self-construction is no mere fine-tuning of the system’s state variables, but an *historical* process in the sense that the system’s very structure changes throughout its lifetime in metastable, path-dependent ways, breaking time symmetries by changing not only its state, but also the parameters and invariants of its entire relational-operational structure. Autonomous systems must necessarily modulate their constraints or boundary conditions, their parametric relations with the environment, and their coupling to other systems, thereby reconfiguring their own phase space in inherently historical ways. Enaction is precisely this ongoing project by which an autopoietic identity reconfigures its operational closure and structural coupling in the interest of its own continuing survival. In this way, the system is able to decouple its actions somewhat from purely environmental determination, and the term “autonomous” describes both those decoupled actions and indeed the system itself.

22. The structural regularities accumulated by adaptive enaction are related in some sense to the meanings by which an autonomous system interprets interactions with its environment, however this notion of meaning contains hidden complexity. By assuming the adaptivity condition, enactivism explicitly emphasises organism–environment interaction, whether sensorimotor, trophic or otherwise. Autonomous systems are therefore *inter-subjective*: they not only interpret their physical and social environment, but also continually shape it and are shaped by it. A simple example is the beaver who niche-constructs its wetlands, which in turn respond with effects that recursively transform the beaver’s own life and that of other species.

23. This essentially plural inter-subjectivity of life and mind is also a source of spontaneity, as organism and environment each adjust to the shifting sands of their co-dependence, making it at least dangerous to think of the meaning of a sign such as “grass” as a mental “*representation*.” The meaningful regularity Grass does not represent some mind-independent grassy-thing. Rather, it enacts the idiosyncratic history of past interactions between a certain family of (long, thin, green) experiences and certain sense-making subjects whose meaning constructions may be as disparate as

those of a grazing cow, an allergy-sufferer or an ant infested by the parasite *Dicrocoelium dendriticum*.

24. Now, Di Paolo's adaptivity condition is a powerful requirement that delivers the enactive accumulation of meaningful regularities through an autonomous system's history. If a system is capable of preferentially accumulating those meanings that are of advantage to its survival, it should indeed enact a set of internal structures that are particularly competent at dealing with a potentially predatory environment. However, adaptivity contains a fatal explanatory gap, since it fails to naturalise the issues of selection, normativity and unitarity entailed in adaptivity itself:

- *How do autonomous systems implement adaptivity?* Evolutionary theory links adaptation to selection in populations, but as Di Paolo (2018: footnote 12) remarks: "The question of whether some form of collectivity is implied in the enactive conception of life is an important one, but remains so far unresolved."
- *By what natural means does adaptivity distinguish normatively between advantageous and disadvantageous regularities?* Plasticity (e.g., Hebbian) rules would already presuppose the adaptive system's ability to distinguish dis-/advantageous variation, while neo-Darwinian selection involves differential reproduction of individuals.
- *Are adaptively autonomous systems causally efficient agents?* Phenomenologically, we usually assume that organisms possess an identity that is able to originate causal effects: *It is I who catches the ball, dammit, not my collective molecular interactions!* Structural coupling delivers me from a passively billiard ball existence of mere reaction to external stimuli, but not from an automated clockwork existence in which internal structures dictate my every action.

25. I claim that we can address these questions by replacing the adaptivity condition in Di Paolo's definition of autonomy by an agency condition, defining enaction, not as *adaptive* autonomy, but as *agential* autonomy. I shall indeed argue that agency *entails* adaptivity in this context. For now, I define:

- A *structure* is a countable set of dynamical relations and operations (*components*). I assume that these components are subject to stochastic variation. Such a structure might for example be an isolated container of gas particles.
- A structure is *operationally closed* if its components reciprocally maintain their collective dynamical state within a stable basin of attraction, thus constituting the structure as a persistent *entity* such as an atom containing subatomic constituents.
- *Structural coupling* of an entity with its environment is constituted by meaningful regularities that enable the entity's dynamical behaviour to avoid being absolutely specified by chance environmental perturbations. This distinguishes the entity as an *identity* such as a refrigerator, whose thermostat maintains its internal temperature against environmental fluctuations.

- An identity possesses *agency* if its meanings are not reducible to any countable set of structural relations and operations – whether internal or external to the agent. This agential identity is then capable of irreducibly originating causal effects on its environment, and I describe it as *autonomous*.

Meaning is intentioned narrative

26. The notion of agency was anathema to cognitive science of the 1980s, which emphasised almost exclusively structural models of cognition as computational operations on relational representations. The prevailing view was that agency is action under the influence of meaning and intentional states, which, as Paul Churchland (1988) eloquently argued, could be the cause of computational operations only if they themselves are prescriptive representational structures.

27. Yet as I noted previously, the notion of meaning as structural representation is problematic if we view it as enacted out of a history of inter-subjective interactions. For example, when we exercise any everyday activity such as walking, we do not engage a single off-the-peg gait operation, but rather interact skilfully and reciprocally with a constantly changing ground surface. The meanings we construct within each modulated contact between ground and feet implement a recursively inter-subjective dynamic to which no computational sequence is adequate.

28. This idea of meaning as inter-subjective dynamic is central to the semiotic account of Charles Sanders Peirce (1932, CP1.320 *et seq.*), who emphasised three skills entailed by organisms' ability to make meaning. *Firstness* is pure sensory quality. For example, if I touch my hand to the firm surface of a closed door, I experience the sensory quality of the surface's resistance. *Secondness* concerns the structures (relations, operations) that arise between qualities. If I press my hand against the door's surface, I experience two sensory qualities: the effort of pressing and the resistance of the surface. These qualities are contingent upon each other, and their relation constitutes the hardness of the surface.

29. Finally, *Thirdness* situates relations within the dynamical context of anticipation. If I use the pressure of my hand to close an open door, the hardness relation gives way to an irreducible dynamic in which the motions of my hand and the door achieve meaning by reference to their anticipated function of closing the door. Peirce proposed that in order for any sensory cue to be a sign, all three of these skills must simultaneously come into play: a *sign* carries a *meaning* for *someone*. Or: The *sign* (First) signifies an *object* (Second) through the mediation of an *interpretant* (Third).

30. I find Peirce's term *interpretant* tricky to pin down. In some texts, it denotes the organism to whom the sign is relevant; in others, it seems more like an interpretive activity. Firstness is our first point of encounter with our environment, Secondness concerns our structural (relational, operational) categorisation of this encounter, and Thirdness involves the living process that links sign to category. I therefore understand

semiosis (the process of meaning-making) in this way: A *sign* evokes a *dynamical* interpretant that chooses some *structural* categorisation of that sign.

31. Despite our experience of semiosis deriving exclusively from living organisms, nothing in this interpretation requires signs to be linguistic, structures to be intentional or dynamics to be mental. For example, Terrence Deacon (2011: 5) describes how bacteria interpret chemical signs in terms of structural constraints that open up dynamical possibilities. Jesper Hoffmeyer (2010) describes semiosis as the “capacity for anticipation” that makes it possible for the living behaviour of formulating useful guesses to evolve from non-living origins. Finally, countless processes in biological development demonstrate that the interpretive linkage between sign and object need not be mental. This realisation led Peirce (1958, CP 7.515) “to the hypothesis that the laws of the universe have been formed under a universal tendency of all things toward generalization and habit-taking.”

32. Bruner (1990) proposed in *Acts of Meaning* an important step toward naturalising this connection between meaning and habit-taking. Bruner understood the goal of cognitive psychology as “to discover and to describe formally the meanings that human beings create” out of their experience (ibid: 2). Throughout his sixty-year professional career, he sought to understand the nature of conceptual thinking, drawing together experimental evidence, reasoning and insights from developmental and educational psychology, to improve the education of underprivileged children. Becoming frustrated with his earlier focus on concepts as categorical relations, he proposed a more organic view.

33. Against the dauntingly structural backdrop of 1980s cognitive models, and prefiguring the dynamical emphasis proposed a year later by Varela, Thompson and Rosch (1991), Bruner sought to reintroduce into cognitive science precisely Peirce’s dynamically anticipative view of meaning. He founded his discussion of meaning not on relations, but on *narrative*, which he characterises (Bruner 1990: 43–50, 77) as:

- giving “voice” to the intentioned perspective of a narrator;
- describing the causally coordinated flow linking events involving participants;
- defining the meaning of events in terms not of participants’ intentions, but of their function within the narrative;
- explicating the mitigating circumstances surrounding conflict-threatening breaches in the canonicity of life.

34. A narrative starts from a structure that is, from some perspective, conflicted, and develops from this structure the coordinated coupling of causes and effects whose meaning derives from their contribution to resolving that conflict. For example, the well-known story of Hänsel and Gretel recounts how their canonical childhood is breached by growing conflict with their step-mother. The story’s narrative develops the causally coordinated consequences of this conflict, leading them to mark a trail through a forest and meet a witch who out of no obvious motivation likes to capture and eat small children. Both forest and witch function to give voice to the narrator’s intention of demonstrating how Hänsel and Gretel mitigate their predicament by learning to rely on

each other's resourcefulness, returning home strengthened by this newfound self-reliance.

35. Situating our interactions with others within a narrative framework helps us to understand their actions in context (Gallagher 2012). A narrative weaves events into a causal unity through time and between participants, its intention endowing these with normativity and meaning. Michele Crossley (2000) discusses how we use narrative to weave our disparate life-roles into a coherent personal identity and self-concept that thereby acquires meaning in relation to others.

36. The participants in these narratives require no beliefs, intentions or desires. Carol Feldman et al. (1990) found no difference in subjects' ability to summarise or recount the details and events of stories that described participants in terms of either intentional states or pure actions. Narratives' capacity to weave together meanings is indeed a very general feature of human cognition: subjects presented with two or three pictures will automatically link these into a story that makes sense of the pictures in terms of the coordination of its participants' interactions (Sarbin 1986). In my own experience, university undergraduates construct such intentioned narratives even when the pictures involve only inanimate objects. Julian Orr (1986) describes how photocopier service technicians solve novel technical problems by recounting "war stories" from daily praxis, then blending these into a single coherent solution narrative through dialogue with colleagues.

37. These qualities of narrative led Bruner to suggest that meaning resides in our capacity to develop stories as interpretants for managing life-disrupting situations:

[Children] produce and comprehend stories, are comforted and alarmed by them, long before they are capable of handling [even simple] logical propositions. Indeed, [...] logical propositions are most easily comprehended by the child when they are imbedded in an ongoing story. [...] One is tempted to ask [...] whether narratives may not also serve as early interpretants for "logical" propositions before the child has the mental equipment to handle them by [...] later-developing logical calculi ... (Bruner 1990: 80)

38. Bruner suggests that meaning relates to the recursive, coordinative dynamics between relational and operational structures in a story. These structures themselves need not possess intention; they acquire meaning by virtue of their function of manifesting the intentions of the story's narrative voice. The charm of this solution is that it builds a bridge between the living and the non-living. If a relation has itself no intrinsic intention, but nevertheless contributes to an overarching extrinsic narrative intention, and is present in the narrative precisely because it contributes to that intention, then the relation acquires meaning through its function of manifesting the narrator's intention.

39. Bruner (1990) proposes that the conceptual components of cognition are not structural categories, but *narratives with meaning and dynamical relevance to the everyday praxis of living*. At each instant of our lives, we – as babies, children and adults – construct meaning from events by situating them within coordinated flows that indicate how they function to mitigate conflicts with our intentions.

40. Bruner's proposal has wider significance, for cognition is not the only meaning-maker: organism life-cycles are biologically instantiated narratives that construct personal meaning out of their entire developmental system of resources (Oyama 2000). For example, an embryonic sea urchin teetering on the brink of survival resolves (or not!) the breach between its current blastocoel state and a maturely developed digestive tract. Filopodia grow within its initially spherical blastocoel, each lacking intention, but anchoring randomly to the opposite wall. Where these filopodia reinforce each other's pull on the wall, they adhere 20 to 50 times more strongly. Their meaning derives not from their individual intentions, but from their function of collectively *choosing* a location for the invagination of the blastocoel wall that will later form the adult sea urchin's mouth.

41. Narrative thus expresses a collective intentionality that underlies both cognitive and biological coping, each thereby offering us by analogy insights into the other. In particular, this analogy suggests very strongly that meaning, intentionality and agency are all collective phenomena that, like the sea urchin's gut, arise co-dependently out of the dynamical coordination of multiple participants. Gerd Müller (1990) pointed out that historically coordinated collectives of cells also generate, in the context of biological development, the novelty of action that we otherwise associate with agency. Consider, for example, Sean Carroll's (2012) parable of the individuals John and Mary, who breakfast daily in the same cafe, but at the different respective times 8:45 and 9:00. For months, they never meet, yet one day John arrives slightly late, causing their paths to cross briefly. In this single encounter, romance is born, changing irrevocably the lives of themselves, their friends and their later children.

42. Equipped with this narrative account of meaning in cognition and biology, we now address the following question: Where does the narrative voice come from? Who breathes intentionality into the narratives of conceptual and embryonic development?

Intention selects narratives

43. Donald Polkinghorne (1988: 150) writes,

“[We] make our existence into a whole by understanding it as an expression of a single unfolding and developing story. We are in the middle of our stories and cannot be sure how they will end; we are constantly having to revise the plot as new events are added to our lives. Self, then, is not a static thing or a substance, but a configuring of personal events into an historical unity.”

Bruner (1990: 116) points out that this incessant need for reconfiguration constrains how we understand our own Self: we define our Self in dialogue with our environment through specific *meanings*, which we in turn construct and utilise through specific *practices*. This duality of meaning and praxis applies to any agent, whether organism, concept or stable object of our perception, and in particular to the notions of *meaning-making narrative* and *self-making agent* that we discussed in preceding sections. A narrative constitutes a dynamically coordinated unity that interprets events in terms of the practical matter of mitigating its narrator's precariousness; an agent constitutes a

causally irreducible identity that reconstructs itself in pursuit of survival. In both cases, what binds together meaning and praxis is the *intention* to stabilise a Self.

44. Comparing these two ideas suggests the hypothesis that an autonomous agent is a narrative instantiated in a collective of participants, and which acts by making meaning from its environment. This implies in turn that the narrative's collective identity must itself somehow provide the irreducible intention that underlies these practical meanings. To argue for this hypothesis, we address in this section two related questions:

- How can an identity be irreducible to its own structure?
- How can such an identity implement intentionality?

45. The first question seems difficult to answer; however, Sober and Wilson (1998) specify quite precisely of the conditions under which a group of participants is irreducible under evolutionary selection. If the survival of individual participants depends upon the presence of some generally available common resource, and if this availability of the resource requires behavioural interaction between those participants, then the survival of the individuals depends upon the survival of their inter-subjective interaction. In other words, the idea of an irreducible group identity makes perfect evolutionary sense whenever inter-participant coordination generates a collective commons that is relevant to the survival of individual participants.

46. To get a feeling for how this plays out in non-evolutionary scenarios, consider the following specific examples of irreducibility arising from a collective commons:

- The genetic material of a bacterium is a countable set of structural relations and operations in the form of condition-action rules: each gene specifies a protein expression rate conditional upon local transcription protein concentrations. These protein expression operations collectively constitute a non-locally, continuously distributed commons of physico-chemical concentration fields that diffuse, condense and adhere to maintain the background preconditions (cell membrane, RNA polymerase, transcription factors) for continued genetic expression.
- When starved, individuals of the slime-mould *D. discoideum* secrete the compound cAMP, forming a commons that coordinates, through quorum sensing, the motions of those individuals to enact a collectively causal fruiting body.
- Hanne De Jaegher and Di Paolo (2007: 497) define *participatory sense-making* as the “coordination of intentional activity in interaction, whereby individual sense-making processes are affected and new domains of social sense-making can be generated that were not available to each individual on her own.” Imagine approaching an unfamiliar person along a corridor and deciding whether to say “Good morning” or “Hi.” Which greeting you choose depends on the developing dynamics of the encounter. Is she looking towards you or away? Does she seem ready to speak? How is she responding to your body language? Your collective, inter-subjectively coordinated dance affects each of you differently, but your final

choices of greeting are definitely not reducible to classifying either of you individually as a Hi- or a Good-morning-person.

47. These examples indicate that irreducible causes are everyday phenomena. None of the coordinated participants need be animate, and the resource they generate can be material substance or mediated choreography. It seems that their coordination is the necessary determinant of irreducibility. It may lead to the organisation of material or energetic resources, but ultimately it suffices if the coordination of multiple participants influences in some way the individual actions of those participants.

48. Relating this analysis to the nature of autonomous agents: If an agent's relational-operational structure generates a coordination whose presence influences the dynamics of those self-same relations and operations, then the behaviour of the collective structure-plus-coordination is irreducible to the behaviour of individual relations and operations.

49. This insight indicates the crucial role of coordination in irreducibility. If we require autonomous systems to be agential (i.e., irreducibly causally), they must be both collective and coordinated, since only collective coordination can deliver irreducibility. Once such inter-participant coordination is reliably present, we may expect this to drive the multi-stable transitions that characterise the complex dynamics of living systems. Which brings us to our second question from the beginning of this section: How might an irreducibly causal agent aspire to intentionality, and how might it align this intentionality with its own self-maintenance?

50. First, what is intentionality? When faced with some novel situation such as news of mass atrocities in Sudan, I must choose how to respond: I *appraise* the situation's relevance to my Self, *intend* some meaningful orientation, and *select* specific motor actions. Appraisal relates to the integrity of my irreducibly coordinated Self, intention concerns my narrative anticipation of events and their consequences, and selection focuses my attention upon one among several options. If Sudan impinges on my coordinated bubble of Selfhood, if I possess some effective narrative of political intervention, and if I wish to be proactive, then I may indeed act.

51. Intention, then, involves irreducible coordination, meaning and closure of options. Now, irreducibly causal coordination requires agents to be durative, or non-instantaneous, since coordination is the intrinsically durative development over time of a constellation of participants. An example system that decides duratively is the Watt steam governor discussed by Tim van Gelder (1995), in which boiler and flywheel participate in an irreducible narrative of mutual co-regulation. In contrast to a digital thermostat, which regulates heating discretely based on temperature data, the Watt governor is an integral controller that continuously time-integrates torque into a speed value, which in turn regulates the boiler pressure that delivers that torque.

52. So intention entails durative coordination of anticipatory meanings. Further, as a living, intentioned agent, I choose by selecting my options for behaviour. That is, my intentions raise the likelihood that I will institute some behaviours, and lower the

likelihood of others. Furthermore, if the participants in my personal narrative are themselves narratives, I do not so much select individual behaviours, but rather the narratives that implement those behaviours in context to maintain, modify or even destroy my identity in accord with some normative perspective. But this is the definition of narrative: the exploration of how to resolve unfamiliar circumstances from some normative perspective.

53. So my intention entails the durative selection between narratives relevant to my own well-being. Moreover, for this selection process to generate genuinely agential actions that originate irreducibly from my identity, it must also contribute to the historicity of my identity. For if it did not, I would then be forever condemned to deciding in the identically same way in all identical situations, and would hence lose my claim to agency. Thus, an agent's irreducibly coordinated selecting must lead to learning: ongoing plastic transformation of the narratives that constitute its identity.

54. To summarise our discussion so far, I claim that an agent is a narrative whose participants are themselves narrative, and their actions acquire meaning through fulfilling certain functions in relation to some normative perspective. The agent constitutes its irreducible identity by making available to those participants a collective coordination conditioned by their individual actions. This coordination supports a causally efficient identity only if it duratively selects structural changes that lead to plastic alterations in the agent's identity and behaviour. Under this account, we can claim that the agent *adapts* if, and only if, these structural changes tend to work in the normative interest of preserving the agent's narrative identity.

Stabilised narratives are intrinsically intended

55. My chain of argument for naturalised autonomy is nearly complete, but lacks one crucial link. The Watt governor is durative and normative, but lacks intrinsic intention: its actions derive all meaning from the extrinsic intention of its designer to regulate the flywheel's speed. By contrast, a living agent's intrinsic intention should strive for goals set by the agent itself: the narrator is the narrative. This invests the agent's actions with value with respect to success or failure in achieving its goals. Di Paolo's adaptivity condition links the normative value of an autonomous system's actions to their capacity for self-production and self-distinction, yet this solution puts the cart before the horse. It posits adaptivity as a process that benefits autopoiesis without explaining how it comes to act predominantly to the advantage of the autonomous system, rather than to its disadvantage.

56. So agency entails adaptivity *provided* the agent's intentions are intrinsic, that is, its constituent narratives are selected with reference to the norm of the agent's own autopoiesis. This is the naturalisation gap in Di Paolo's account of enactive autonomy: alignment cannot be declared by fiat, but must arise as a natural consequence of the structural changes that create and produce the agent. Now, we have already assumed

that structures are susceptible to stochastic variation, and Goodwin (1994: 59) pointed out that the essence of self-maintenance as a variational constraint is the natural, entirely non-teleological process of *dynamical stabilisation* (henceforth abbreviated to *stabilisation*) of this stochastic variation.

57. I offer here two illustrations of my understanding of stabilisation, based on personal conversations with Goodwin between 1986 and 2001. Firstly, imagine a seashell settling on a sandy seabed. Initially, it rests precariously upon the sand, wafted by passing currents in coordinated motion with the sand underneath it. Both shell and sand grains move randomly, but whereas the shell's motion is non-local, constrained by its own solid form, the motion of the grains is individually localised. The shell's dynamics coordinate the grains' random movements, while they in turn constrain the shell's dynamics. Whenever the shell shifts position, this creates interstices between itself and the seabed, into which grains drift opportunistically, packing and buttressing the entire shell's new orientation. Occasionally, a catastrophic toppling of grains forms a hollow into which the shell collapses, this new shell-sand configuration again creating opportunities for consolidation by randomly drifting grains. *Stabilisation* refers to this ratchet-like relaxation into shell-sand stability, leading us naturally to observe stabler configurations more frequently than less stable ones.

58. Secondly, I view stabilisation as evolving out of Goodwin's collaboration in the 1950s with Conrad Waddington, who introduced the terms *genetic assimilation* and *canalisation* to account for the following experimental observations. Waddington (1959) exposed *Drosophila* eggs to dangerously high concentrations of salt (sodium chloride). When the eggs hatched, they displayed a wide variety of unexpected, mostly fatal, phenotypes, including one with enlarged anal papillae (EAP): organs associated with balancing osmosis. He selected these phenotypes for breeding, exposing the next generation to similarly high salt concentrations; this increased the frequency of EAP phenotypes in the experimental population. After several generations of this selection procedure, the EAP phenotype became fixed in the fruit-fly population and continued to manifest even in individuals whom Waddington exposed to much lower, more typical salt concentrations.

59. Waddington described these results in the following way. High salt concentration is an environmental determinant that plastically induces the trait EAP in some individuals, whereupon (artificial) selection increases the frequency in the fruit-fly population of genes that functionally correspond to this plastic trait. Subsequently, genetic assimilation fixes, or canalises, these genes in the population, making the trait stable against variations in the original inducing determinant of high salt concentration. James Baldwin (1896) had previously proposed such an evolutionary mechanism, which also underlies Mary Jane West-Eberhard's (2003) proposal of *developmental plasticity* as a primary driver of evolutionary adaptation.

60. Let us consider three different ways in which we might naturalise Waddington's account of this experiment from the respective perspectives of *DNA structure*, *niche coordination* or *organism life-cycle narrative*. From a *DNA-centric* perspective,

Waddington's experiment involves discrete, stochastically varying relations between DNA base pairs (adjacency, folding, chromatin structure). These structures are entirely passive, relegating any potential agency to non-local fields (electric, concentrations of salt, polymerase and signalling proteins) in the DNA's cellular and extracellular context. These fields coordinate developmental processes such as the expression and mutated replication of the DNA structures, enabling the frequency of successful genetic structures to rise to the point where they swamp out the effect of other, less successful, genetic variants. As it becomes increasingly difficult for new variants to assert themselves in competition with currently successful ones, the level of genetic variability in the population decreases.

61. Whereas neo-Darwinism focuses on changing population genetic frequencies, Goodwin suggests we instead focus on this concomitant rise in the population's variational stability. From the DNA-centric perspective, these two approaches have equivalent explanatory power. Both explain how a high salt environment might condition a rise in stability of those DNA mutations that by chance happen to develop the EAP phenotype, yet neither explains how fruit-flies initially respond to high salt concentrations by rapidly diversifying the phenotypic variants, later generalising the chance EAP solution from high- to low-salt concentrations.

62. Switching to a *niche-centric* perspective, we can view Waddington's experiment as simulating a niche: a quasi-closed collection of non-locally interacting physical fields (electromagnetic, chemical, elastic), whose reactive, diffusive and conductive properties mediate between local operations due to fruit-flies or to Waddington himself as experimenter. Waddington manipulates a particularly simple coordination field – a non-local salt concentration, holding it dynamically stable, and thereby promoting variational stability of the fruit-fly phenotypes by permitting their genotypes to drift opportunistically into stable states. These phenotypes may coincidentally influence their internal local salt concentration, and their variational stability will increase if these phenotypes can themselves enhance the dynamical stability of that internal concentration. That is, stabilisation offers a payoff for constraining the internal salt concentration of the individual fruit-fly phenotype.

63. These two perspectives become especially interesting when we combine them into a third perspective centred on organisms' *developmental narrative*. In this perspective, the life-cycle of the individual fruit-fly constructs itself out of structural DNA relations and coordinating fields of the surrounding niche. By consuming resources, excreting, and building bodies, nests or a chemical micro-environment, the fertilised egg bootstraps from these fields an *umwelt* of personal relevance, collaterally influencing the development and evolution of other organisms. As Susan Oyama (2000) argues incisively, a fruit-fly is not the passive product of pre-specified genetic or environmental factors, but the *meaningful narrative of niche-constructing a material body out of structures and coordination*.

64. This narrative perspective makes perfect sense of Waddington's experiment. The fruit-fly population is itself a narrative of individuals possessing varying genetic and

phenotypic structures coordinated by developmental and environmental flows. Initially, salt concentration is statically low, permitting the population's genetic structures to drift into a variationally stable state that provides initial conditions for stable development in a low-salt environment. From the niche-centric perspective, we expect the fruit-flies' developmental narrative to have stabilised itself, through osmotic or molecular chaperoning mechanisms, against small changes in salt concentration, but not against Waddington's dangerously high concentrations. Under these survival-threatening conditions, we might expect such stabilising mechanisms to unravel, permitting precisely the proliferation of phenotypic variants that Waddington observed.

65. This destabilisation will also likely make the developing phenotypes more susceptible to changes in environmental fields such as salt concentration, diversifying phenotypes even further. If some structural variant coincides by chance with some reliably stable coordination such as the salt concentration in such a way as fortuitously to develop the EAP phenotype, we expect from the DNA-centric perspective that this will increase the variational stability of this variant within the population. This stabilisation moves the population historically from the low-salt to a high-salt developmental attractor, and this attractor is conditioned not only by the new salt concentration, but also by the durative transition from one to the other. Because of this, it seems plausible that there will stabilise itself within the population a genetic structure that abstracts the ability to switch rapidly between *both* developmental attractors.

66. The above account is speculative – I shall test it in the following section. However, it illustrates Goodwin's thinking about stabilisation, which he considered a naturalisation of evolutionary fitness. The slogan "*Survival of the fittest!*" suffers from two shortcomings: it applies only to the population genetics of reproducing individuals, and it obscures its own tautological use of the term "fitness," for which reproductive success, or "survival," is the only practical measure. Goodwin effectively replaces this tautology by the rather less sexy slogan: "*Narrative is the mutual stabilisation of structural variation and coordinative dynamics!*"

67. Stabilisation naturalises Jakob von Uexküll's proposal of a general normative directedness in nature that highlights the role of meaning and creativity in the emergence of life from matter. Von Uexküll pointed to the fit between a butterfly's proboscis and the tubular bloom from which it feeds as a structural abstraction from an ongoing narrative in which each participant acquires meaning through interaction with the other. The mutual coordination of their semiotic encounter downwardly constrains its participants' genetic components, and these genetic components emerge through their influence on that coordination. He dismissed discussions of purpose from biology, yet "what remains uncontested is the presence of a [meaning] rule in living Nature, which reveals itself even in the mechanical processes of the organism" (Uexküll 1926: 270).

68. I claim that stabilisation is the single constitutive mechanism of living systems. Stochastically varying structural components engender by chance a local region of stable coordination, which in turn has the potential to attenuate the level of variation in

those structural components. If structure and coordination happen to stabilise each other reciprocally in this way, it opens a co-dependently arising bubble of narrative that stabilises itself against degradation from its external environment. Goodwin (2009) believed that stability is the only possible natural norm, all other norms deriving ultimately from the intentions of participating agents, which themselves, I claim, are grounded in narrative stabilisation.

69. Narrative agents make choices when they respond to some event by living a narrative whose coordination constrains the distribution of all possible actions onto a smaller subset. Equally, an agent chooses to learn when it assigns to some event a narrative meaning whose coordination constrains the distribution of all possible structural variations onto a smaller subset. I define the following terms for narrative stabilisation as a naturalised basis for both of these constraining processes:

- A *dynamical* system possesses discrete, local structure and continuous, non-local coordination (state variables/fields). The structure determines the parameters and relations that condition the coordination's time-flow (typically modelled by differential equations).
- A *narrative* is a dynamical system whose structure is subject to stochastic variation, and whose coordination constrains that variation.
- *Narrative stabilisation* is the process by which a narrative's coordination constrains its structural variation in such a way that structure and coordination converge to a (meta-)stable attractor.
- An *agent* is a narratively stabilised narrative. The agent's stable structure determines its *behaviour*, its coordination defines its intrinsic *intentions*, and its own narrative stabilisation defines its *norms*.

70. Under these definitions, the agency condition for autonomy entails adaptivity if narrative agents are capable of enacting their history in structure – that is, if they *learn*. The next step in my naturalisation argument is to demonstrate this ability of narratively stabilised agents to acquire meanings that are adaptive in the sense of increasing those agents' stability in a changing environment.

Stabilised narratives are autonomous

71. In §2 above I formulated three goals: (A) characterise life; (B) analyse this characterisation into necessary constraints upon living systems; and (C) demonstrate that some unequivocally natural implementation satisfies these constraints. In the preceding, I have (A) characterised life as agential autonomy: an autonomous agent originates behaviour that is unitarily and causally irreducible to its external context and/or internal constitution. I have further (B) analysed agential autonomy into the constraint that autonomous agents arise and survive through narrative stabilisation. That is, an agent bootstraps its unitary self as a stabilised collective of sub-narratives that collectively condition a common dynamical coordination field, which in turn, through

natural mechanisms of stabilisation, constrains stochastic variations of these sub-narratives. The collective influence of sub-narratives on the coordination field implements *behaviour*; the selective effect of coordinative stabilisation on structural variation implements *adaptation*; the recursive interplay of behaviour and adaptation implements autonomous *choice*.

72. There remains just one final gap in this naturalisation argument. I have suggested strongly that the single normative criterion of being stabilised is sufficient to support adaptation in narratively stabilised agents. In order to demonstrate that this is in fact the case, I must present (C) a natural example of a narratively stabilised agent that learns, that is, the agent constructs new structures that enhance its own stability. Such an agent would fulfil Di Paolo's adaptivity criterion and would therefore constitute an enactively autonomous agent. To this end, I present in this section a computational, and hence (C) unequivocally natural, implementation WattWorld that exhibits this ability. WattWorld is a very simple Julia-language implementation of narrative stabilisation, and its required credentials for life-like behaviour are that ...

- a. this narrative stabilisation is capable, out of a population of extrinsically intentioned narratives, of constructing an agent – that is, a narrative that possesses stable, coordinative, intrinsic intention;
- b. this agent exhibits ontogenic plasticity: learning to behave skilfully to implement this intrinsic intention;
- c. the agent exhibits phylogenic adaptation: learning to generalise this skilful behaviour across contexts.

73. The following analysis of WattWorld is purely qualitative, leaving a more quantitative analysis for another occasion. This analysis will demonstrate that WattWorld (a) spontaneously constructs a narrative, whose stability defines its own normative intention (regimes 1–3 below). This narrative (b) defines itself ontogenically by developing skilful coordinative behaviours to maintain its stability in the face of environmental perturbation (regime 4). Finally, the narrative (c) learns to generalise these stability-maintaining behaviours phylogenically across contexts (regimes 5–6).

74. Motivated by the previously mentioned Watt governor, WattWorld draws upon Andrew Watson and James Lovelock's (1983) Daisyworld: a simulated world populated by differentially reproducing integral controllers that individually raise or lower their common resource of planetary temperature. Similarly, WattWorld treats boiler pressure as a common resource ($R \in [0.0, 10.0]$), and contains a population of five (unintentioned) *player* narratives p that are randomly initialised, and which raise or lower the value of R , depending on their mutable, individual *saturation* attribute ($K_p \in D_K \equiv [-2.0, -0.1] \cup [0.1, 2.0]$). WattWorld's coordination variables comprise the players' individual *activation* attributes ($a_p \in (0, 3.0]$) and the continuously varying global resource R . The players' structure conditions the time-flow of the coordination variables R and a_p according to the following differential equations.

75. The resource level R is subject to a time-dependent external feed $F(t)$, exponential depletion $-\beta_R R$ (where β_R is a constant 1.0) and regulatory control $-a_p K_p$ by each of the five players p :

$$\frac{dR}{dt} = F(t) - \beta_R R - \sum_p a_p K_p$$

Positive values of K_p lower R , while negative K_p values raise R . The activation a_p of player p varies dependent upon R according to the following differential equations:

$$\frac{da_p}{dt} = a_p \left(h \left(\sum_q \alpha_q |K_q|, -1.0 \right) \cdot h(R, K_p) - \beta_a \right)$$

$$h(x, K) \equiv \begin{cases} \frac{x}{x + K}, & (K > 0) \\ 1 - h(x, |K|), & (K < 0) \end{cases}$$

The sum in the first equation extends over the respective products of the activation α_q and the magnitude $|K_q|$ of the saturation parameter for each player q (independently of p); $h(x, K)$ is a Hill saturation function of x , with saturation constant K and cooperativity 1.0. The time development of a_p is proportional to a_p , and contains two terms: the first dependent upon $h(R, K_p)$, and the second an exponential depletion term $-\beta_a a_p$, where β_R is constant (0.1). Collectively, these differential equations describe an integral-rein controller (Saunders et al. 1998) of the commons R , in which players regulate R in proportion to their individual activation a_p and saturation K_p . R in turn conditions an increase or decrease in their activations a_p .

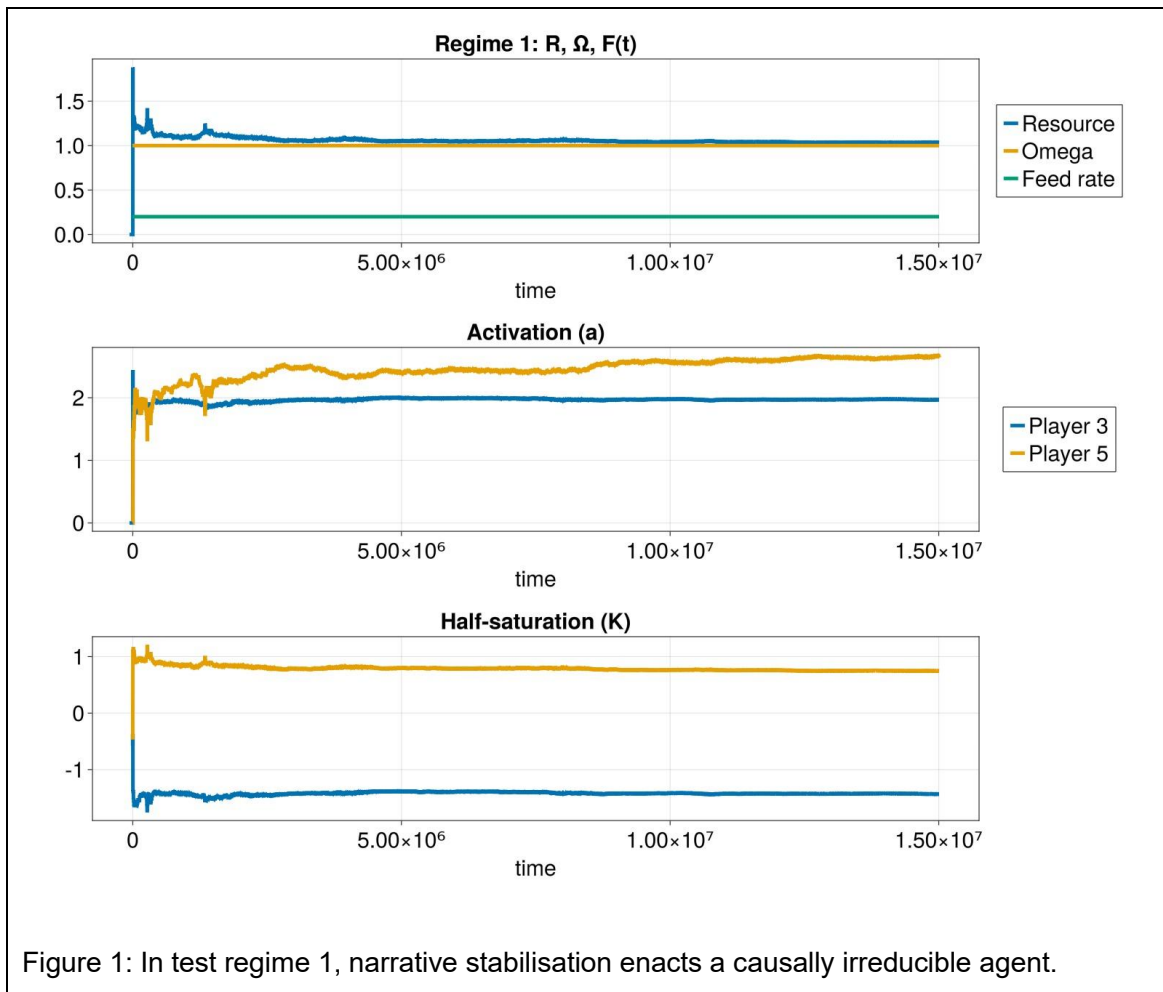
76. Collective integral-rein control of R can only occur if the players' structural parameters K take appropriate values, and it is our hope that WattWorld's stabilisation will locate these values. At each simulation time-step, WattWorld first uses Runge-Kutta-2 integration to calculate the next incremental change of coordination variables due to the above differential equations. It then subjects the structural parameters K of all players to stochastic variation δK , which, purely for convenience, wraps across all boundaries of the saturation constant domain D_K . This variation is constrained by a very simple stabilisation rule, namely, that δK_p (for all p) approaches zero as R approaches a predefined value Ω (the *context* of the simulation), according to the following rule:

$$\delta K_p \leq 0.05 \min(1.0, |R - \Omega|^{2.5})$$

Regime 1: Narrative stabilisation enacts agency

77. Figure 1 shows the results of the first WattWorld test regime, in which the external feed $F(t) \equiv 0.2$ and context value $\Omega = 1.0$ are both held constant. In the top graph, we see that WattWorld almost perfectly regulates R to Ω . It achieves this (bottom graph) by stabilising the saturation constants of players 3 and 5 to $K_3 \approx -1.5$ and $K_5 \approx 0.7$. Thus, WattWorld stabilises into a coordinated, integral-rein narrative in which players 3 and 5

control R by respectively upregulating and downregulating its value. Times are given in simulation time-steps, and only players satisfying $a_p > 0.01$ are displayed.



78. That this narrative controls R is not surprising; perfect regulation is the nature of integral controllers. What regime 1 demonstrates is firstly, that narrative stabilisation *stabilises*. WattWorld follows only one imperative: that its entire structure becomes sticky, in the sense that all variation approaches zero, as R approaches Ω . It possesses no arbitrary plasticity or selection rules that distribute reproductive rewards to specific players. That stickiness alone can stabilise a functioning narrative is not obvious.

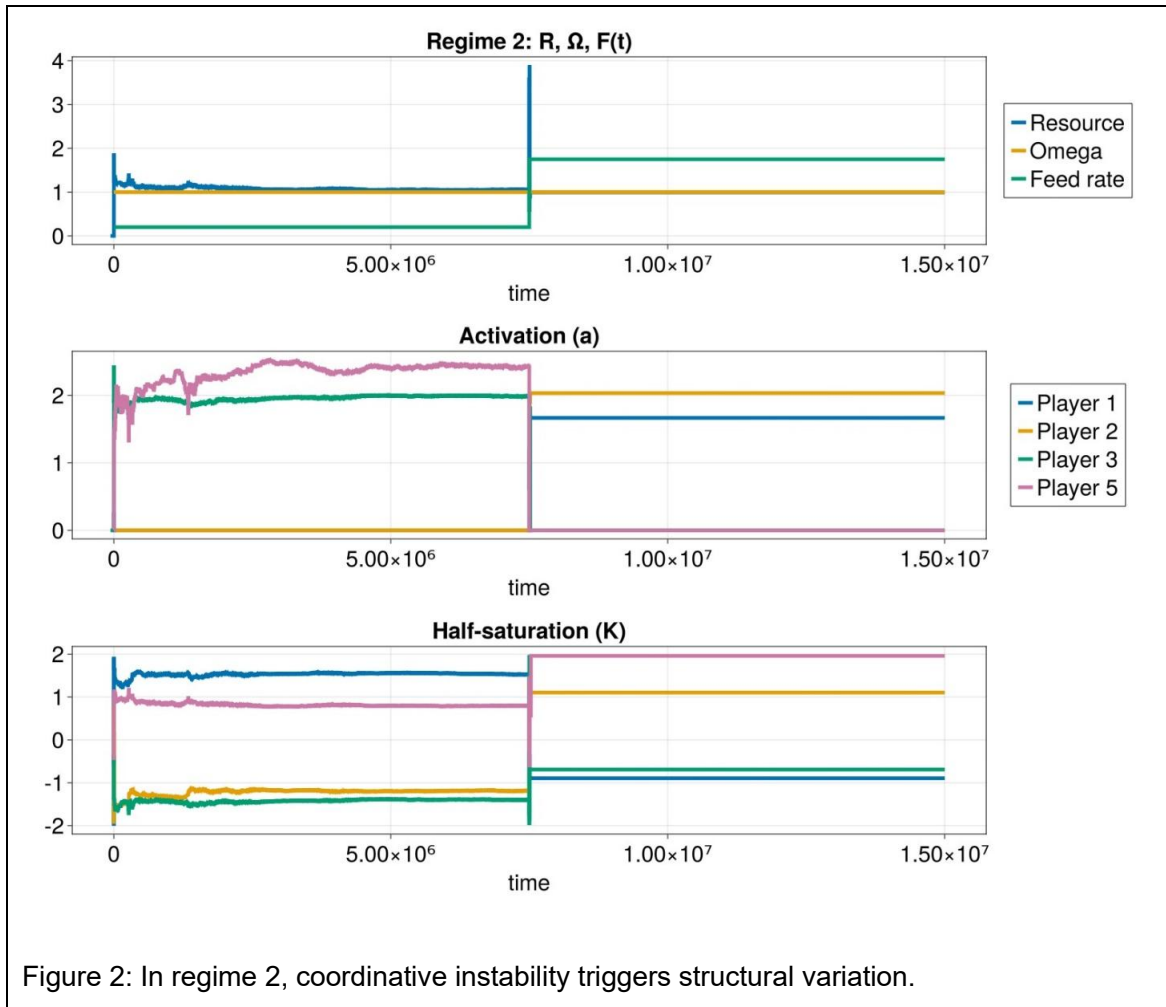
79. Secondly, this stabilisation enacts an agent constituted by the combined narrative of players 3 and 5, since this narrative is both causally irreducible and intentional. It is *irreducible*, since it is only able to control R as a causal entirety – neither player alone is capable of achieving this. And it is *intentional* since the central cause of its stable presence is its ability to mitigate its own dissolution that would result from being unable to control R in this way. The narrative both contributes to its own stability and exists stably because of that contribution, and thus coordinates its interaction with WattWorld intentionally with respect to the norm of stability.

80. I am indebted to an anonymous reviewer for pointing out that the collective agent is enacted not only by players 3 and 5, but indeed by all five players. For the low activation of the remaining players is also a collective choice that contributes to agency. Thus, although players with low activation do not influence the agent's control

behaviour, they do influence how this controller will adapt in response to changes in context, as we will see in later regimes.

Regime 2: Coordinative instability triggers structural instability

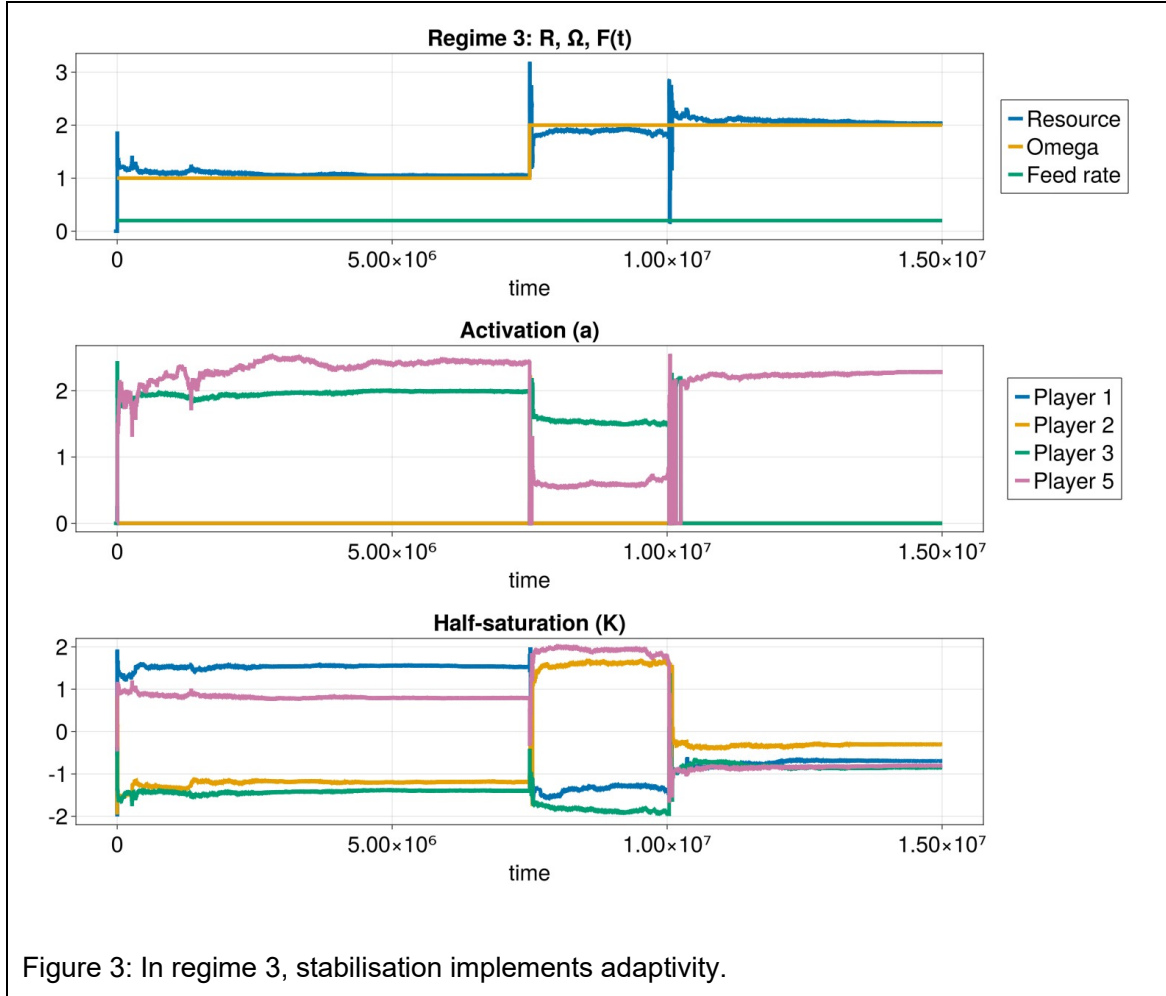
81. Figure 2 shows the results of regime 2, in which the external feed $F(t)$ takes a single step from 0.2 to 1.75 after 7.5×10^6 time-steps, while context Ω remains constant at 1.0. Given that the first phase of this regime enacts an integral-rein agent comprising players 3 and 5, we might expect this agent to cope well with the feed step, since integral-rein controllers respond accurately to feed changes by recalibrating the activations a_p .



82. Yet surprisingly, the step triggers a transient jolt in the value of R (not visible at this graphical resolution), which in turn destabilises WattWorld's structure, permitting drastic changes in K_1 and K_2 and the emergence of a new, stabler, agent comprising players 1 and 2. We see that WattWorld supplements integral-rein control's familiar plastic coordination with the possibility of deeper structural change. As with Waddington's fruit-flies, ontogenic (coordinative) instability triggers phylogenetic (structural) instability, accelerating the search for ever more stable configurations.

Regime 3: Structural stabilisation implements adaptivity

83. In regime 3, the feed $F(t) \equiv 0.2$ stays constant, while the context Ω takes a single step from 1.0 to 2.0 after 7.5×10^6 time-steps; this is a much more difficult problem for integral-rein agents, which control R by trapping its value at the intersection between a falling ($K < 0$) and a rising ($K > 0$) Hill function. The agent can adjust to moderate changes in feed without changing the K structure of the Hill functions, by coordinating activations to preserve the intersection value, but changing Ω requires the agent to adapt the K_p to an entirely new intersection point.



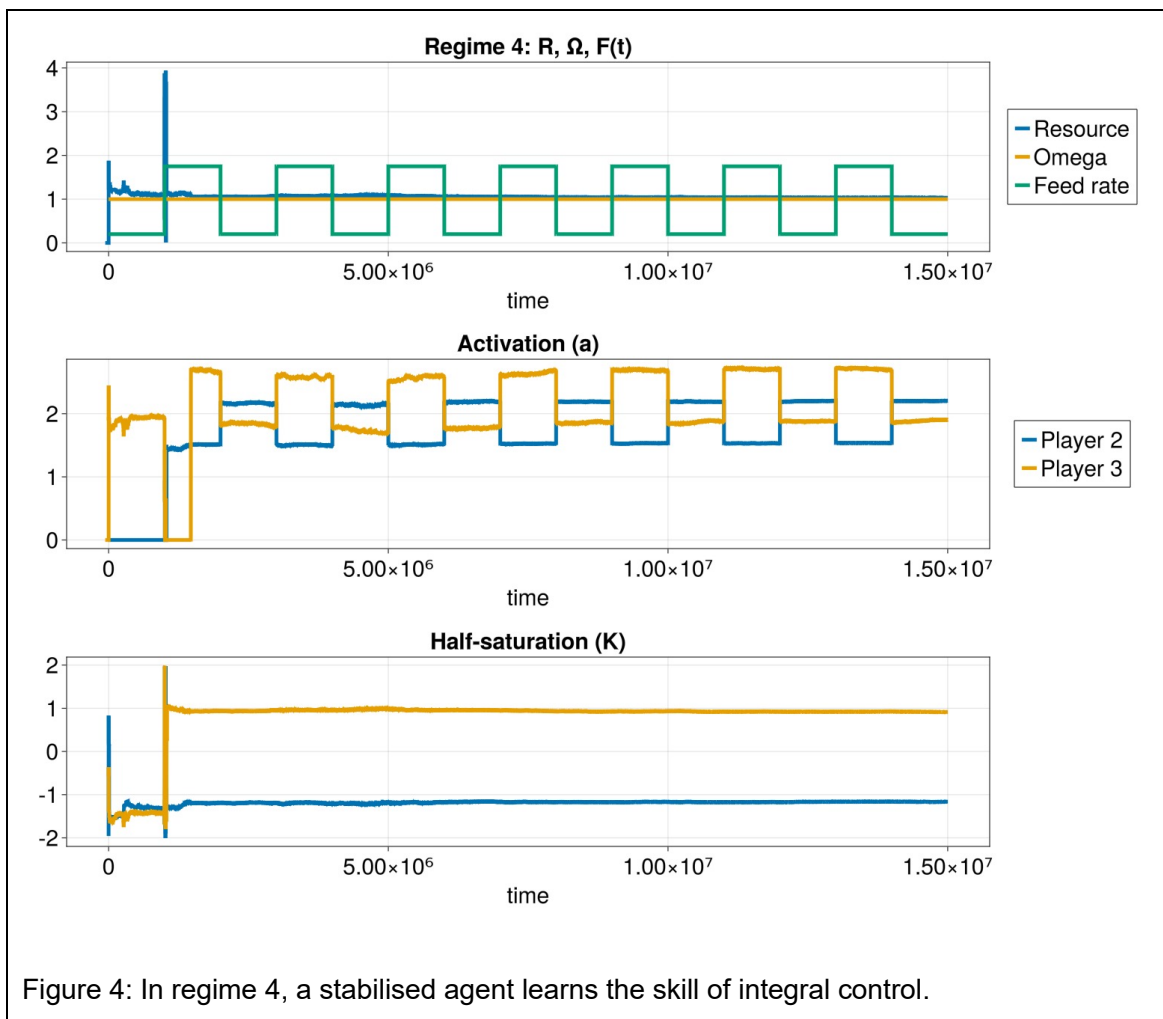
84. In the top graph of figure 3, we see that narrative stabilisation copes with the step in Ω , though less cleanly than in regime 2, because of the need to adapt the K_p ; prior to the step, this run is identical to regime 2. After the step, players 3 and 5 initially adjust their activations to cope tolerably with the change in Ω , but this imperfect regulation frees WattWorld's structure to vary, until at around 10^7 time-steps, K_3 hits the lower boundary (-2.0) of the saturation constant domain D_K , wrapping between widely differing values in D_K and precipitating abrupt, complex change. Out of this tumult, player 5 emerges as the single player regulating R with $K \approx -0.8$.

85. Regime 3 demonstrates that structural stabilisation is able to adapt regulation to changes in context. Yet the question arises, whether this regulation by a single player fulfils the irreducibility condition for agency: does this configuration of WattWorld still

constitute an agent? Indeed, it does, since this (admittedly poor) regulation of R relies not only upon player 5 being highly activated, but also upon other players maintaining low activation. In this regime, it is no longer so clear precisely which players constitute the adapting agent; particularly during the complex phase after 10^7 time-steps, all players participate in developing the later configuration.

Regime 4: Coordination implements skilful action

86. Regime 4 demonstrates agents' ability to learn the classic skill of integral controllers: integral control. Here, $\Omega = 1.0$ stays constant, while $F(t)$ flips periodically between 0.2 and 1.75. Within one cycle of the feed function, players 2 and 3 coordinate to control R , and from 8×10^6 time-steps onwards, this coordination copes stably and seamlessly with changes in the value of F . This fine-tuning of skill is achieved through structural stabilisation of K_2 and K_3 .



Regime 5: Structural stabilisation optimises coordination

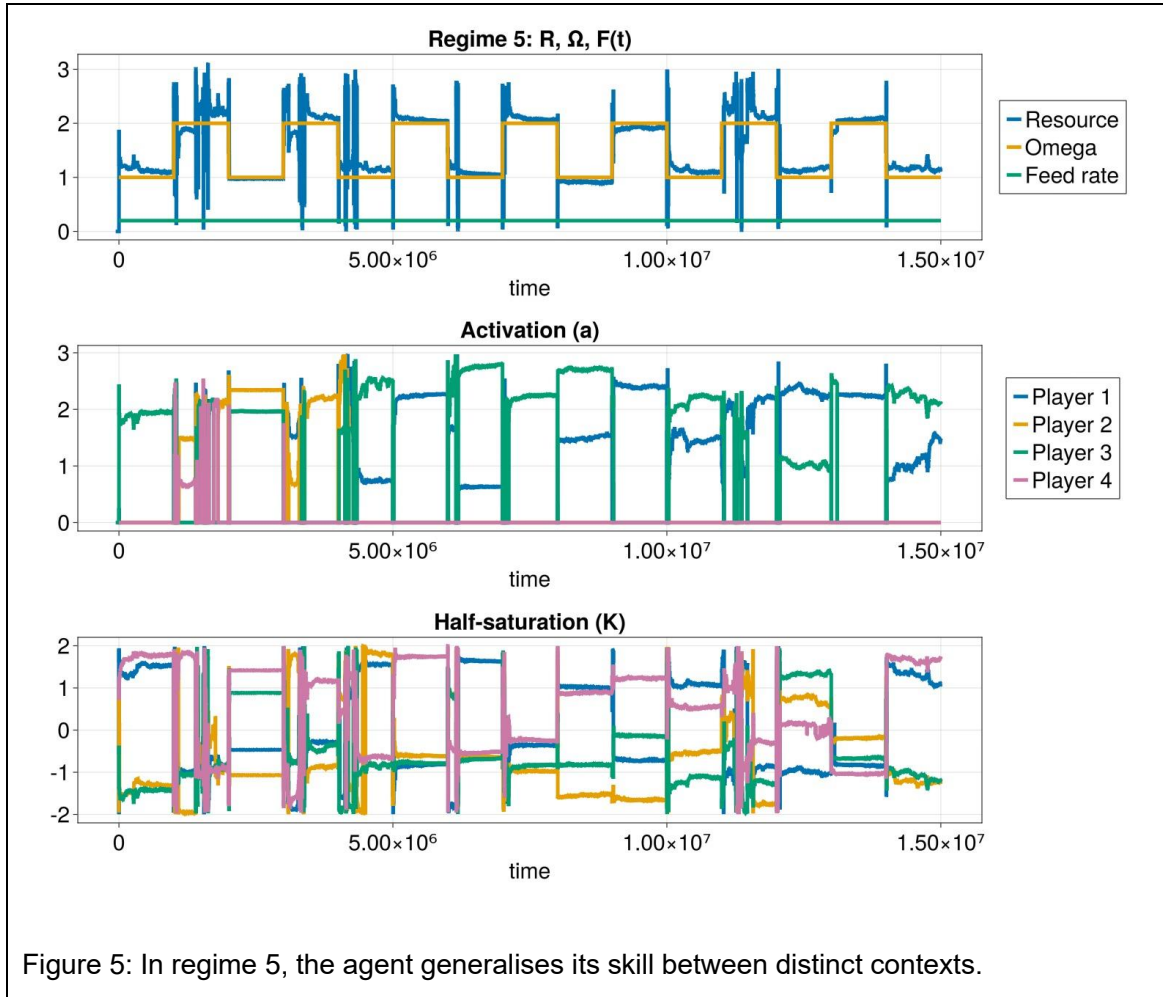


Figure 5: In regime 5, the agent generalises its skill between distinct contexts.

87. Figure 5 illustrates a harder learning task: flipping periodically between the contexts $\Omega = 1.0$ and $\Omega = 2.0$ while $F(t) \equiv 0.2$ stays constant. The top graph demonstrates how WattWorld adapts to reduce the disruption of stability that results from Ω -switching. This skill is briefly disrupted at around 1.12×10^7 time-steps when K_1 and K_3 thrash between positive and negative values, but then settles into a narrative that avoids the need for rapid structural change by re-coordinating a_2 between contexts. Although such apparently stable configurations can repeatedly collapse and reform during a simulation run, their reconstruction is nevertheless a reliable feature of WattWorld.

Regime 6: Agency entails adaptivity and abstraction

88. In regime 6, $F(t)$ fluctuates on a rapid cycle between the values 0.2 and 1.75, while the context Ω fluctuates on a slower cycle between 1.0 and 2.0. Consequently, WattWorld must learn to regulate R against fluctuations in F , while at the same time generalising this skill across contexts.

89. As time progresses along the top graph, we see that both frequency and amplitude of the discrepancy between R and Ω diminish over time, as WattWorld seeks a flexible structural basis for the coordinative skill of periodic regulation. Prior to 4×10^6 time-steps, player 4 plays a significant role (see centre graph), but this solution regulates

poorly at high values of Ω , leading to thrashing of K_3 between negative and positive values until at around 4×10^6 time-steps, player 4's activation falls toward zero. From this point onward, players 2 and 3 collaborate to achieve good regulation over phases of unchanging context Ω , and continue to improve the quality of regulation across periodically switching contexts.

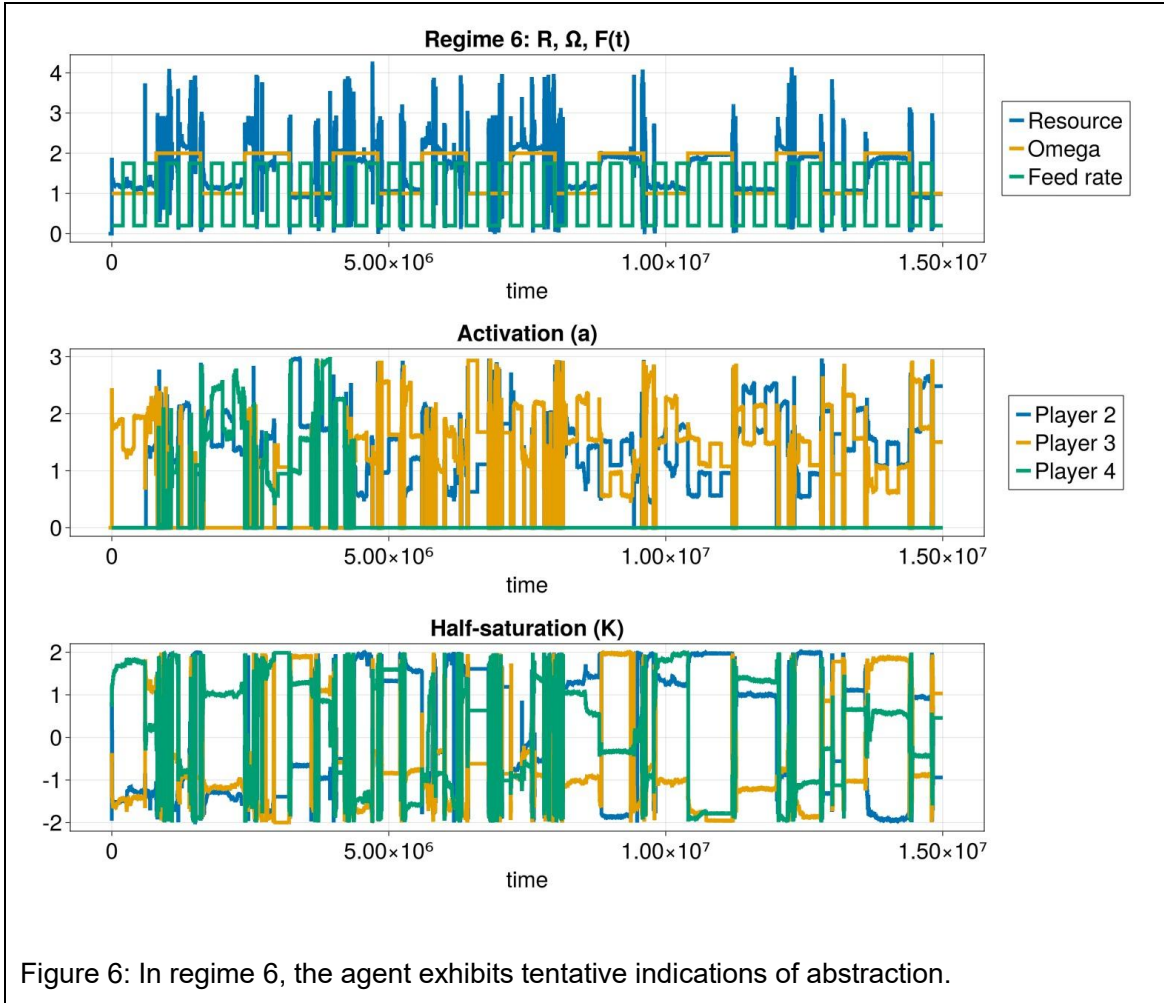


Figure 6: In regime 6, the agent exhibits tentative indications of abstraction.

90. Regime 6 demonstrates firstly that a narratively stabilised agent is genuinely adaptive: it develops an ontogenic, coordinative skill, then adapts its structure to generalise this skill phylogenically across contexts. Secondly, WattWorld exhibits here a central mechanism of this adaptation proposed by Kant: the abstracting of universalisable relations. For compare the second half of the centre and lower graphs in Figure 6, where the coordination a_p follows the rapid periodic fluctuation of F , whereas the structure K_p follows rather the slow fluctuation of Ω . WattWorld records structurally those aspects of its coordinative experience that are stably generalisable across contexts.

91. Again, it would be misleading to suggest that this abstraction of coordinative generalisability into structural stability is persistently stable; indeed, if we extend the duration of the simulation, these abstracted structures repeatedly collapse and reform. However, their construction is highly reliable, exhibiting ever increasing stability – and hence autonomy – as the simulation time progresses.

Conclusion

92. Abramsky et al. (2025) suggest that representation, fundamental to living systems, may constitute a fundamental natural principle analogous to those of physics, while semiotics views representation as an irreducible triad: sign-interpretant-category. I contend in this paper that interpretants are stabilised, or equivalently, intrinsically intentioned, narratives. A narrative *N* is an irreducible exploratory intention to resolve some unfamiliar situation by reference to the normative perspective of some narrator. If this narrator is simply *N* itself, then *N* is its own intrinsic intention. I argue that if *N* arises through natural stabilisation, it is necessarily intrinsically intentioned. A computer simulation of natural processes then demonstrates that narrative stabilisation suffices to implement adaptive, and hence also enactive, autonomy. Narrative stabilisation, based purely on the mutual co-conditioning of structural variation and coordinative dynamics, constitutes precisely such a natural, life-enabling principle as that suggested by Abramsky et al.

93. Although autonomous action is the central pillar of Kant's account of living cognition, he applies the term autonomy not to individuals choosing how to act, but rather to actions arising from principles that are both universally stable across the physical distinctions between experiences and lawful across the personal distinctions between individuals. This definition requires that any cognitive system is necessarily embodied as a unitary collective of peers that collectively *choose* in the sense of forming a *sensus communis* (Kant 1996c, §40, 5:294).

94. From these definitions, Kant derived his famous Categorical Imperative as the moral requirement upon autonomously acting members to maintain the society that supports them:

“All maxims as proceeding from our own law-making ought to harmonise with a possible kingdom of ends as a kingdom of nature.” (Kant 1996a: 4:436)

95. Three steps lead from Kant's kingdom of ends to narrative. First, Di Paolo (and Varela before him) extended the definition of autonomy to describe the agents that can entertain autonomous principles; second, Bruner broadened the definition of “principle” to the intentioned term “narrative”; third, Goodwin emphasised the recursively-coordinated stability that characterises selection of narratives. Narratively stabilised agents are causally irreducible and are adaptive to the extent that they stabilise the autonomy of the ecological niches (including the cellular society of an organism) on which their own autonomy depends. Ecological stability therefore constitutes the intrinsically inter-subjective norm for stochastic variation of the agent's structure.

96. If we understand knowledge as arising from this recursively inter-subjective experience of living, we as organisms are in the same cognitive boat as my granddaughters tentatively toddling upright, or Charles Darwin discerning correlations amongst Galapagos finches: how do we abstract reliably useful knowledge from this unremittingly reciprocal participation? I claim that narrative stabilisation abstracts bubbles of stability from lived experience, stabilising them against stochastic variation

into a narrative self that exhibits stable universality across life situations. Organisms and conceptual principles are narratives containing sub-narratives, and are themselves participants in more inclusive narratives. The defining criterion for the autonomy of such narratives at any level is whether they constitute themselves through narrative stabilisation, which in turn entails Di Paolo's adaptivity criterion for enactive autonomy.

97. At the population level, this recursively coordinative nature of living systems echoes Augustine of Hippo's imperative to, "Love, and do what you will". For I suggest that to experience loving relationship is to become aware of participating meaningfully within wider autonomous narratives that afford a stable home for our own quirky, autonomous individuality. In becoming aware of this, we recognise that our actions both influence these wider narratives, and are in turn constrained by them. Out of this realisation follows all adaptive, moral, loving action.

References

- Abramsky, S., Banzhaf, W., Caves, L.S.D., Levin, M., Machado, P., Ofria, C., Stepney, S. & White, R. (2025). Open questions about time and self-reference in living systems. arXiv:2508.11423v1.
- Baldwin J. M. (1896) A new factor in evolution. *American Naturalist* 30: 441–451.
- Bruner J. S. (1990) *Acts of meaning*. Harvard University Press, Cambridge MA.
- Carroll S. B. (2012) *Endless forms most beautiful*. Quercus, London.
- Churchland P. M. (1988) The ontological status of intentional states: Nailing folk psychology to its porch. *Behavioral and Brain Sciences* 11: 507–508.
- Crossley M. L. (2000) *Introducing narrative psychology: Self, trauma and the construction of meaning*. Open University Press, Buckingham.
- De Jaegher H. & Di Paolo E. (2007) Participatory sense-making. *Phenomenology and the Cognitive Sciences* 6: 485–507. <https://cepa.info/2387>
- Deacon T. W. (2011) *Incomplete nature: How mind emerged from matter*. W. W. Norton, New York.
- Di Paolo E. (2018) The enactive conception of life. In: Newen A., de Bruin L. & Gallagher S. (eds.) *The Oxford handbook of 4E cognition: Embodied, embedded, enactive, and extended*. Oxford University Press, Oxford UK: 71–94. <https://cepa.info/5608>
- Feldman C. F., Bruner J., Renderer B. & Spitzer S. (1990) Narrative comprehension. In: Britton B. K. & Pellegrini A. D. (eds.) *Narrative thought and narrative language*. Lawrence Erlbaum, Mahwah NJ: 1–78.
- Gallagher S. (2012) Neurons, neonates and narrative: From empathic resonance to empathic understanding. In: Foelen A., Lüdtke U. M., Racine T. P. & Zlatev J. (eds.) *Moving ourselves, moving others*. John Benjamins, Amsterdam: 165–196.

- Goodwin B. C. (1994) *How the leopard changed its spots: The evolution of complexity*. Scribner, New York.
- Goodwin B. C. (2009) Genetic epistemology and constructionist biology. *Biological Theory* 4(2): 115–124. Originally published 1982. <https://cepa.info/4651>
- Hoffmeyer J. (2010) A biosemiotics approach to the question of meaning. *Zygon* 45(2): 367–390.
- Kant I. (1996a) *Groundwork of the metaphysics of morals*. Translated by Mary J. Gregor in *Practical Philosophy*. Cambridge University Press, Cambridge. German original published in 1785.
- Kant I. (1996b) *Critique of practical reason*. Translated by Mary J. Gregor in *Practical Philosophy*. Cambridge University Press, Cambridge. German original published in 1788.
- Kant I. (1996c) *Critique of judgement*. Translated by Mary J. Gregor in *Practical Philosophy*. Cambridge University Press, Cambridge. German original published in 1790.
- Kelso J. A. S. & Engstrøm D. A. (2006) *The complementary nature*. MIT Press, Cambridge MA.
- Maturana H. R. & Varela F. J. (1980) *Autopoiesis and cognition: The realization of the living*. Reidel, Dordrecht.
- Maturana H. R. & Varela F. J. (1987) *The tree of knowledge*. Shambhala, Boston.
- Müller G. B. (1990) Developmental mechanisms at the origin of morphological novelty. In: Nitechi M. H. (ed.) *Evolutionary innovations*. Chicago University Press, Chicago: 99–130.
- Orr J. E. (1986) Narratives at work: Story telling as cooperative diagnostic activity. In: *Proceedings of the 1986 ACM conference on Computer-supported cooperative work*. ACM Press, New York: 62–72. <https://doi.org/10.1145/637069.637077>
- Oyama S. (2000) *Cycles of contingency*. Duke University Press, Durham NC.
- Peirce C. S. (1932) *Collected papers of Charles Sanders Peirce, Volumes I and II: Principles of philosophy and elements of logic*. Edited by Charles Hartshorne and Paul Weiss. Harvard University Press, Cambridge MA.
- Peirce C. S. (1958) *Collected papers of Charles Sanders Peirce. Volumes VII and VIII: Science and philosophy and reviews, correspondence and bibliography*. Edited by Arthur W. Burks. Harvard University Press, Cambridge MA.
- Polkinghorne D. E. (1988) *Narrative knowing and the human sciences*. SUNY Press, New York.
- Sarbin T. G. (1986) *Narrative psychology: The storied nature of human conduct*. Praeger, New York.
- Saunders P. T., Koeslag J. H. & Wessels J. A. (1998) Integral rein control in physiology. *Journal of Theoretical Biology* 194(2): 163–173.

- Smith L. B. & Thelen, E. (1993) A dynamic systems approach to development: Applications. MIT Press, Cambridge MA.
- Sober E. & Wilson D. S. (1998) *Unto others*. Harvard University Press, Cambridge MA.
- Uexküll J. von (1926) *Theoretical biology*. Harcourt, Brace & Co, New York. German original published 1920 by Verlag von Gebrüder Paetel, Berlin.
- Van Gelder T. (1995) What might cognition be if not computation? *Journal of Philosophy* 91(7): 345–381.
- Varela F. J. (2000) *El fenómeno de la vida [The phenomenon of life]*. J. C. Sáez, Santiago de Chile.
- Varela F. J., Thompson E. & Rosch E. (1991) *The embodied mind*. MIT Press, Cambridge MA.
- Waddington C. H. (1959) Canalization of development and genetic assimilation of acquired characters. *Nature* 183: 1654–1655.
- Watson A. J. & Lovelock J. E. (1983) Biological homeostasis of the global environment: the parable of Daisyworld. *Tellus* 35B: 284–289.
- West-Eberhard M. J. (2003) *Developmental plasticity and evolution*. Oxford University, New York.