

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Governance Cognitive Systems
Unifying print-principles framework for studying Governance
Level Cognitive Systems, its implications and future research
applications

verfasst von | submitted by

Artur Serrasqueiro Deus Dionisio

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 013

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Joint-Degree Middle European
interdisc. master prog. in Cognitive Science

Betreut von | Supervisor:

Seth Frey Ph.D.

Abstract	4
German Abstract	4
Acknowledgements	6
Dedications	6
Introduction	7
Research aims and Questions	7
Methodology	8
Goal	8
Part I	
First principles	10
Preamble	10
Chapter 1	
Information	14
1.1 (Information as) Decreases in alphabets of states	14
1.1.1 When I learned about probability spaces	15
The sample space	16
The collection of events	16
1.1.2 "Information" Desambiguation	16
1.1.3 System, revisited & defined	18
1.1.4 Information across levels	19
1.2 Systems as historical restraints-cascades	21
1.2.1 Shared sub-paths and the role of assembly bottlenecks	23
1.2.2 Assembly across scales and substrates	24
1.3 Singularity-nodes as sites of level-formation	26
1.3.1 Persistence (without teleology)	26
1.3.2 Formalising viability	27
Phase transitions in constrained ensembles	27
Information bottleneck bifurcations	28
Two complementary formalisations, one criteria	29
1.3.3 Singularity-nodes	29
1.3.4 Selection as path-reweighting	31
1.3.5 Next stop, the reductionist substrate	31
Chapter 2	
Non-equilibrium Integrity, Biology and Cognition	35
2.1 Acquired tastes: from constraints to operations (Cognition as persistence in non-equilibrium)	35
2.2 Motivated collaborators: integrity all the way down	42
2.3 From collaborator to generator	48
2.4 A Kaleidoscope of Selves	52
2.5 Cognition, Governance or principles of operating integrity?	62
2.6 Nested Interests: Alignment Across Dependent Scales	74
2.7 Functionalism across substrates	77
Part II	

En route	85
Chapter 3	
Humans in Assembly	85
3.1 The Trajectory Continues: Distributed Cognition and the Scoping Problem	85
3.2 Assembly Post Organism	94
3.2.1 Culture in the Assembly Continuum	94
3.2.2 The Abstraction Bottleneck	96
3.2.3 Functional Role Emergence Through Assembly	99
3.2.4 Cognitive Glue: The Binding Mechanism	100
3.2.5 Assembly Past the Human Object	105
3.2.6 Language as key Assembly Bottleneck in Cultural Index	108
Chapter 4	
Stable Post-Human Assemblies	111
4.1 Stable recipes for iteration	111
4.1.1 Constitutional Code: Referring to Classical Framings	111
4.1.2 Constitution, Functional Role Crystallisation, Assembly and Integrity	112
4.2 From Humans to Generators: Cognition at the Governance Scale	118
4.3 Part the whole	127
4.4 Recap: Deriving Governance from first principles	134
Part III	
Filling the Gap	145
Chapter 5 — What the state of the art is missing	145
5.1 Limitations in Cognitive Science, Biology and Assembly Theory	145
5.1.1 A Pioneering, Yet (self)Constrained, Foundation	145
5.1.2 Scale-Free Yet Not In Assembly's first principles	149
5.1.3 Analytical, yet Without Ontological Commitment	150
5.1.4 The Persistent Anchoring to Individual Cognition	152
5.2 Limitations in Political Science	154
5.2.1 The Inheritance	154
5.2.2 The Cartesian Architecture Made Visible	157
5.2.3 The Empirical Costs	158
5.2.4 The reckoning	160
Chapter 6 — Final Thoughts on Adding to the State of the Art	161
6.1 What Becomes Visible	161
6.1.1 Governance morphology is governance action	161
6.1.2 A non-ideological taxonomy of Governance Systems	162
6.1.3 The dependency on lower-level dynamic drives	163
6.1.4 Policy as morphological active inference	164
6.2 Hard-Protocol Design Implications	165
6.2.1 Feedback integration	165
6.2.2 Memory consolidation	166
6.2.3 Attention and metacognition	167
6.2.4 Cross-Markov-blanket generative model coordination	167
6.2.5 A broader claim	168
6.3 Dynamic-Drive Design Implications	168

6.3.1 Policy that leverages multiscale competence	169
6.3.2 Surfacing latent drives	170
6.4 Future Research	172
6.4.1 For Cognitive Science	172
6.4.2 For Political Science	174
6.4.3 For Governance Cognitive System science	175
6.4.4 For the AI Era	177
6.5 Conclusion	179
Glossary	181
References	186

Abstract

This thesis grew out of a long-standing curiosity about how institutions, constitutions and governance systems come to behave *as if* they had goals of their own, irreducible to any single human who composes them. Trained in both Law and Cognitive Science, I found that neither field, on its own, offered ground on which such questions could even be properly posed. The first step was therefore to build a common analytical substrate from which both disciplines could be re-derived from shared first principles.

I commit to a monist, physicalist, functionalist epistemology, and select a precise informational definition as the reductionist substrate from which I derive: the criteria for systemhood (positive total correlation); the grammar of level nesting (recursive coarse-graining); the accumulation of constraints across time (Assembly Theory's bottlenecks and path-dependent memory); and the configurations at which new levels consolidate (singularity-nodes). Once non-equilibrium thresholds are crossed, I treat Cognition as what any assembly does to actively maintain its integrity through prediction-error minimisation, with the same architecture recurring across cellular, organismal, cultural and governance scales.

Reading governance through this lens, I reframe Constitutions as recursive assembly memory, that is, as code encoding both its first-order constraints and the procedures for its own iteration, and Governance Cognitive Systems as higher-order Markov blankets whose generative models, variable sensitivities and pathologies become objects of empirical study. I introduce the distinction between hard protocols, dynamic drives and compatible assembly index as co-determinants of system viability and close by sketching the research lines this synthesis opens. Among them, the conditions under which constitutional code remains viable on a given cultural-scaffolding assembly index and the embodied costs that governance trajectories impose on the humans who constitute them.

Keywords: information theory; Assembly Theory; Active Inference; Basal Cognition; Markov Blankets; Constitutional Theory; Governance; multiscale competency; functionalism.

German Abstract

Die vorliegende Arbeit ist aus einer lange gehegten Neugier darüber hervorgegangen, wie Institutionen, Verfassungen und Governance-Systeme sich so verhalten, als hätten sie eigene Ziele, die auf keinen einzelnen sie konstituierenden Menschen zurückführbar sind. Mit einem Ausbildungshintergrund sowohl in der Rechtswissenschaft als auch in der Kognitionswissenschaft stellte ich fest, dass keines der beiden Fächer für sich genommen einen Boden bietet, auf dem solche Fragen überhaupt angemessen aufgeworfen werden können. Der erste Schritt bestand daher darin, ein gemeinsames analytisches Substrat zu errichten, von dem aus beide Disziplinen aus geteilten ersten Prinzipien neu hergeleitet werden können.

Ich verpflichte mich auf eine monistische, physikalistische und funktionalistische Epistemologie und lege eine präzise informationelle Definition als reduktionistisches Substrat zugrunde, aus dem ich Folgendes herleite: die Kriterien für Systemhaftigkeit (positive Gesamtkorrelation); die Grammatik der Ebenenverschachtelung (rekursive Vergrößerung); die zeitliche Akkumulation von Beschränkungen (die Engpässe und das pfadabhängige Gedächtnis der Assembly Theory); sowie die Konfigurationen, an denen sich neue Ebenen konsolidieren (Singularitätsknoten). Sobald Nichtgleichgewichtsschwellen überschritten sind, fasse ich Kognition als das auf, was jede Assembly tut, um ihre Integrität durch die Minimierung von Vorhersagefehlern aktiv aufrechtzuerhalten – wobei dieselbe Architektur über zelluläre, organismische, kulturelle und Governance-Skalen hinweg wiederkehrt.

Liest man Governance durch diese Linse, so deute ich Verfassungen als rekursives Assembly-Gedächtnis um, das heißt als Code, der zugleich seine Beschränkungen erster Ordnung und die Verfahren seiner eigenen Iteration kodiert; kognitive Governance-Systeme erscheinen entsprechend als Markov-Decken höherer Ordnung, deren generative Modelle, Variablensensitivitäten und Pathologien zu Gegenständen empirischer Untersuchung werden. Ich führe die Unterscheidung zwischen harten Protokollen, dynamischen Antrieben und einem kompatiblen Assembly-Index als mitbestimmende Faktoren der Systemviabilität ein und schließe mit einer Skizze derjenigen Forschungslinien, die diese Synthese eröffnet. Darunter die Bedingungen, unter denen Verfassungscode auf einem gegebenen Assembly-Index kultureller Gerüstbildung lebensfähig bleibt, sowie die verkörpert Kosten, die Governance-Trajektorien jenen Menschen auferlegen, die sie konstituieren.

Schlüsselwörter: Informationstheorie; Assembly Theory; Active Inference; Basale Kognition; Markov-Decken; Verfassungstheorie; Governance; mehrskalige Kompetenz; Funktionalismus.

Acknowledgements

Thank you to those who directly impacted this thesis:

- Seth Frey for supervising this thesis and for showing me that Chekhov's gun really ought to be fired. I truly hope to count on your collaboration and insights for many years to come
- Markus F. Peschl for the community you've built and for the passion of your teachings
- John Clippinger and Michael Zargham for the kind support, availability and championing of the ideas which lead to this thesis
- Carolina Czizek for the friendship and seeing this thesis' light when it was only a slight sparkle

Dedications

To:

- My Mother¹ who gave me life and taught me the noble lesson of keeping brave perseverance throughout adversity
- My Father² who saved my life and taught me the brightness of pure curiosity and adventurous deep wonder
- My brother³, who I admire and for whom I strive to one day be the best possible irmão.
- My partner⁴ who inspires me to become the best version of myself and teaches me the value of unassuming, natural profundity

¹ Maria Francisca Dionísio

² Orlando Deus Dionísio

³ António Deus Dionísio

⁴ Inês Silvestre, aka Cosmos

Introduction

Research aims and Questions

This thesis constitutes the vindication of both my system's thinking approach to Law School⁵ and my Protocol oriented approach to Cognitive Science. In it lies the deep seeded motivations that lead me to an unusual set of academic backgrounds, in my curiosity driven urge to understand the mechanisms whereby society organises itself. It has long made intuitive sense to me to regard *The Polis* as the first *artificial intelligence*⁶ and *the hive* as an organism of its own. What stood underneath these unexplored intuitions was the articulated insight into how emergent systems give rise to, pursue and order goals which are irreducible to their constituent parts.

The questions that brought me here had been with me long before I could phrase them properly in the problem-space above. Trying to make sense of the world is in a big part trying to figure how it is that *institutions* (composed of, run by, and dependent upon individual humans) come to behave *as if* they had their own goals, irreducible and often conflicting to those of any one participant⁷. While making sense of cultural change and its epistemic implications is in no small measure trying to understand why do *Constitutions*⁸ and laws persist across the turnover of those who interpret and enforce them, while the individuals themselves often pay personal costs to sustain them⁹? Being a deterministic monist¹⁰, I knew that pursuing these questions encompassed understanding what is the relationship between a cell maintaining its integrity within a tissue, an organ maintaining its function within an organism, and a citizen maintaining their role within a State.

These were not questions I could find adequately addressed at the intersection of the two disciplines I had trained in. Cognitive Science had grown comfortable extending the unit of analysis to artefacts, tools and culture yet stopped short of treating *Governance* as cognitive in its own right, while Political Science kept out of focus the major advances in theories of goal-directed and information processing yield across Cognitive Disciplines¹¹. Neither field, on its own, offered even a common angle on which these questions could even be properly posed in a manner that would bridge their respective areas of expertise. So I knew the absolute first step was the attempt to construct that common analytics substrate that would bring both disciplines to a level of analysis in which their concepts could be re-derived from common first principles and made to converge on a shared object of study.

⁵ I naively started my bachelor in Law believing I would be learning something akin to societal design, the study of the imposition of restrictions and incentives that lead to societal outcomes. The field and course is actually much more casuistic, though I did study Economics of Law and Philosophy of Law, which presented me to a lot of the themes I came back to in this thesis. Plus I am grateful for the practical and heuristics skills my law degree endowed me with

⁶ Although the making of this thesis made me understand the concept of *artificial* as superficial, and I hereby stand by the claim that there isn't something like *artificial* intelligence, simply *intelligence*, which can be instantiated in a variety of types and embodied across multiple substrates

⁷ Nobody *wants* pollution, global warming, systeming inequality, zero sum atomic war protocols

⁸ And their nations and institutions, however what I come to understand *Constitutions* as is later developed in this thesis

⁹ As I am writing this section a Ukrainian colleague of mine pause work and is on a training camp in the context of the Russian invasion

¹⁰ Yes, I had no free will in writing this

¹¹ My diagnosis in this thesis, derived from a distrust of Positivism, was that Law and Political Science inherited an implicit *Cartesianism* that kept institutions and the people who enact them on different ontological sides of an analytical line.

Methodology

This thesis is theoretical, foundational, and exploratory in a first-principles sense. I begin by securing an epistemological footing that is explicitly monist and reductionist, functionalist, and materialist, the 'Isms' that furnish a minimal common lens through which cognitive and political concepts can commonly travel free of *baggage*. I select *Information* as the level-independent substrate expressed as constraints reweighting lower-level probability distributions while preserving physical continuity across scales. Then, from this substrate, I derive the criteria for systemhood, the grammar of level nesting, the temporal dimension of constraint accumulation and the locations at which new levels of description consolidate. To pursue my analysis focusing on configurations that persist and propagate, generating incremental complexity.

With this scaffolding in place, I then assess the literature in both fields through the same lens, treating all relevant variables across disciplines by what they do in *state spaces*. The conclusions are then all derived from these angles. With *Governance* understood as assembled systems actively pursuing integrity and their *Constitutions* as stable, self-referential recipes for iteration that instantiate persistence. The sequence is always epistemologically grounded on the parameterisation of phenomena into a common substrate, from which further concepts are critically read and interpreted to provide a synthesis between Cognitive and Political sciences based on their functional role within this common *reductionist substrate*. This is the method that both organises the structure of the thesis and makes explicit at each derivation the consequences of reducing cognitive and political principles to *information processing*.

Goal

With this thesis I aim to provide a foundational synthesis between Cognitive and Political Science, by framing Governance Structures as Cognitive Systems, to both yield a common conceptual framework to these scholar traditions¹² and lay the ground for a common object of study between the fields. It is oriented at deepening the understanding of societal Cognition elicited through Governance Protocols and aimed at future practical (e.g. betterment of policy making) and theoretical (e.g. study of cognitive properties within different protocol architectures) scientific applications at a newly found interception between these fields.

The arc of the thesis builds the philosophical footing required for cross-scale and cross-domain analysis without dualism, establishes what Agents are, so as to understand Cognition as operations that maintain system level integrity, traces how Distributed Cognition becomes abstracted through functional roles, and from these a synthesis is consolidated. Thus, along the way, it derives implications, names variables that permit comparison and provides examples and illustrations to corroborate the arguments.

Despite major advances in intelligent driven systems, institutional design (which aims at creating and maintaining intelligent *organs*) has not kept pace. This thesis, treating Governance as Cognitive System, renders these institutional *organs* and *constitutions* empirically tractable, opening the door towards a scientific programme in a which mechanisms can be specified, performance assessed with universal informational criteria, and architectures of policy and oversight prototyped, tested and iteratively improved.

¹² Structuring their concepts from a minimal common branch

My motivation is practical and clear. If Governance Systems are recognised as Cognitive Systems, we can evaluate designs by their informational performance, how well they maintain integrity under perturbation, how quickly they learn from error, how effectively they distribute attention and how transparently they generate and track goals. This allows policy to be specified as a cognitive architecture, taking into consideration all of the variables cognitive disciplines take into consideration when designing *intelligent solutions* and leveraging them into Policy Making. It also clarifies what must be protected if we wish to preserve democratic Cognition, not only rights as declarative statements, but the integrity of the feedback channels, memory stores and attention regimes that make those rights operative in practice.

Part I

First principles

Preamble

This thesis, being an exercise in scoping Cognition so as to encompass processes traditionally analysed in Political Science, faces many of the challenges shared by interdisciplinary efforts¹³. My attempt at mitigation these is finding a common ground upon which foundational arguments can be laid. It is thus not possible to build bridges sustained by semiotic or hermeneutical reasoning extrapolated from *Cognitive* and *Political* symbols, since doing so would be to rely on representations that stand on different realms of contextual, historical and situational schools of thought. Rather, my aim is to reexamine the information underlying concepts of both disciplines, tracing them back while anchoring them in first-principles, so as to escape the boundaries of legacy definitions and the limitations of siloed traditions.

This approach is informed by two beliefs. First, that definitions are insufficient to contain the essence of that which they aim to entail (Wittgenstein, 2009, §66-§67); and second, that knowledge resides not within disciplinary containment but rather in a place of holistic disregard for academic borders (Wilson, 1998). The consequence of such beliefs is holding that — at the core of currently siloed representations separately held by *Political* and *Cognitive* Sciences — there are underlying parameters which, through their architecture or manifestation, express properties capable of synthesising a new shared representation between the two fields.

The first step on my journey towards this thesis is finding such eligible parameters, defining them and determining their scope. In this exercise I begin by trying to isolate a common minimal denominator with resource to an exercise in reductionism.

Beyond a methodological approach — of *general* explainability from the analysis of common *components*¹⁴ — reductionism is a philosophical stance not only on science but reality itself, positing that only what can be revealed through analytical processes is real¹⁵. It operates under the assumption that complexity can be explained by the same universal principles that govern its simplest components¹⁶. And thus, critically for synthesising Cognitive and Political science, that legacy disciplinary terminology can be ultimately subsumed by common underlying principles¹⁷. While searching for a common minimal denominator between Cognitive and Political Science, reductionism offers a straightforward, systematic pathway. Identifying and breaking down concepts and terminology from both disciplines into fundamental patterns, constraints and processes that undergird both fields.

¹³ Such as barriers in jargon and conflicting terminology, paradigmatic problem space differences and the methodological dissimilarities stemming from their disparity, assessment and critical thinking difficulties due to lack of established metrics or difference in approaching criteria, or even deeper foundational Philosophical Divergences (steaming, for example, from positivists versus relativists stances or other such epistemological attitudes) (Lach, 2014), (Daniel et al., 2022).

¹⁴ "Reductionism is the idea that all of the complex and apparently disparate things we observe (...) can be explained in terms of universal principles governing their common ultimate constituents" (Nagel, 1998, p. 3).

¹⁵ "Reductionism affirms that only what is revealed in an analysis is real. Analysis is also affirmed as ultimately the only legitimate approach for our understanding of the world and as the goal that all science should have" (Jones, 2013, p. 19)

¹⁶ "The idea is that there exists, in principle, a theory of everything, in the form of a theory governing the one thing or few things of which everything else consists." (Nagel, 1998, p. 3)

¹⁷ "The explanatory thesis is that everything [underlying by me, to express the natural inclusion of behaviour] that happens can be given an ultimate explanation in terms of the laws governing those elements." (Nagel, 1998, p. 3)

This is critical not only to assure that two disciplines loaded with psychological and social concepts are anchored to a monist and empiricist approach (treating more *ethereally* perceived *mental* or *cultural* processes as phenomena founded in the same natural and physicalist reality, and not as separate ontological domains), but also to avoid non-reductionist explanatory routes that lack a tangent and falsifiable route for synthesis (such as Jung's Collective Unconscious, Hegel's World Spirit and other psychological and political theoretic constructs that aren't reducible to an elemental analytical substrate).

However, given that this approach is dependent upon selecting a ground floor from which its principles are derived — the elemental building block of a given reductionist approach, from which analysis and explanations are derived, what could be named its *Reductionist Substrate* — there are huge downstream implications stemming from what *reductionist substrate* is selected. Historically, the insight of analysing reality through its smallest, most fundamental parts, traces back to ancient Greece, with Democritus — *atomism's* epitome — asserting that reality consists of eternal, indivisible *atomon* as the fundamental basis of reality (Taylor, 1999). Yet the *Reductionist Substrates* selected in *explanation building* have historically shifted as our understanding of nature evolved and across different scientific requirements (as per the scale, type and order of a given analysis, or phenomenon, of interest).

Allow me to showcase two prominent examples from The Scientific Revolution to illustrate *Reductionist Substrate* selection, its malleability and potential implications. Francis Bacon (*Novum Organum*, 1620) broke down analysis to a process of induction reduced to observable properties¹⁸, deeming all which is not observable and directly measurable as constructs which ought to be reduced to phenomena available to empirical individual observations¹⁹. Bacon's *Reductionist Subtract* wasn't *Constitutive* — that is — it doesn't make the ontological claim that the universe is composed by the common element *observable properties*²⁰, but it is nevertheless *Explanatory*, subsumed being the assumption that all events can be explained *in function* of those reductions²¹. By contrast, Descartes' binary reductionist subtract, however, is based on the ontological claim that reality is composed of two fundamentally different categories: *Res Extensa*, encompassing all material objects²² (though, fundamentally, rooted in geometry²³); and *Res Cogitans*,

¹⁸ "Man, as the minister and interpreter of nature, does and understands as much as his observations on the order of nature (...) permit him, and neither knows nor is capable of more." Bacon, on *Novum Organum*, Aphorism I, affirming the primacy of observable phenomena; "Our method (...) consists in determining the degrees of certainty, while we, as it were, restore the senses to their former rank (...) and attain our end... from the first actual perceptions of the senses themselves." *Novum Organum*, Aphorisms—Book I, Preface, illustrating the reductionism towards directly observable (perceived) phenomena as the building block from which knowledge is constructed

¹⁹ "We have no sound notions either in logic or physics; substance, quality, action, passion, and existence are not clear notions; much less weight, levity, density, tenuity, moisture, dryness..." (Francis Bacon, *Novum Organum*, Aphorism XV)

²⁰ In fact, it is fair to speculate that Bacon would regard Democritus' atomism as an "Idol of the Theatre": "The subtlety of nature is far beyond that of sense or of the understanding: so that the specious meditations, speculations, and theories of mankind are but a kind of insanity." (*Novum Organum*, Aphorism X). Plus, while exposing the systematic biases that cloud human understanding (referred to as "idols"): "Lastly, there are idols which have crept into men's minds from the various dogmas of peculiar systems of philosophy (...) we denominate idols of the theatre: for we regard all the systems of philosophy hitherto received or imagined, as so many plays brought out and performed, creating fictitious and theatrical worlds." (*Novum Organum*, Aphorism XLIV)

²¹ Further sections explain how *Reductionist Substrates* can be explanatory and relational while still being anchored in first-principles

²² "Descartes' natural philosophy begins with his conception of body. For Descartes, of course, extension is the essence of body or corporeal substance. Or, to use the technical terminology that Descartes adopted in the *Principles*, extension is the principal attribute of corporeal substance." (Cottingham, 1992, p.292); "The notion of extension is so closely bound to the notion of corporeal substance that, for Descartes, we cannot comprehend the notion of this substance apart from its principal attribute" (Cottingham, 1992, p.293)

²³ In fact, an argument can be made that Descartes' true reductionist substrate — common to both *Res Extensa* and *Res Cogitans* — are axiomatic, clear and distinct ideas. Take Cottingham (1992) perspective on how *Res Extensa* can be reduced to Geometry, and Tho (2020) on how this can further be distilled to axiomatic truths: "Cartesian bodies are just the objects of geometry made real, purely geometrical objects that exist outside of the minds that conceive them" (Cottingham, 1992, p.294); "Descartes grounds knowledge on the clearness and distinctness of ideas, of which geometrical certainty is a paradigmatic

encompassing mental substance²⁴. A reader familiar with history of science will be quick to note that although both such authors defer to *elemental building blocks* from which study and explanations are derived — as a method of decomposing and pursuing a problem space (what I am shortening to *Reductionist Substrate*) — by virtue of the different *Substrates* selected, legacies of Dualism vs Monism, Idealism vs Materialism, Rationalism vs Empiricism were pursued. The choice matters because it predetermines what one looks at and what counts as an explanation. This ought to inflate the importance of carefully navigating the problem of selecting a *Reductionist Substrate*, as well as its type²⁵.

Threading such a task with great care — as choice of *reductionist substrate* is the foundational pillar from which the epistemological scaffolds are built and first-principles are derived — I have started to probe my *Reductionist Substrate* selection problem space by studying critiques of *Reductionism* overall.

Regardless of being prone to shifts in substrate, Reductionism has been subject to criticism both in its *Constitutive* and *Explanatory* forms, the nature of which I've agglomerated to three general types: Epistemological Criticism; Ontological Criticism and Downward Causation Criticism. While the first I am promptly dismissing²⁶, the other two have deeply informed my approach.

The Ontological Criticism posits that "*Some physical phenomena, even though they can be explained in terms of principles that fit their specific features, simply do not have an explanation at the ultimate level*" (Nagel, 1998, p. 5). If we think about the phenomena of *domestication of animals* through the lens of particle physics as a reductionist substrate, we might be able to extrapolate higher order DNA and genetic differentiation through time of lineages of matter agglomerates, but it is arguable if *domestication* can really exist at that level²⁷. Without taking into consideration the higher order notions of morphological

case. Hence complex forms of knowledge are certain if they are rooted in the solid foundation of fundamental axiom-like clear and distinct Ideas" (Tho, 2020), "Descartes grounds knowledge on the clearness and distinctness of ideas, of which geometrical certainty is a paradigmatic case. Hence complex forms of knowledge are certain if they are rooted in the solid foundation of fundamental axiom-like clear and distinct Ideas" (Tho, 2020). Though this — and its implications to the mind-body problem — constitute a whole potential other thesis' subject, it will not be discussed here, as the example above suffices to the current thesis at hand, to illustrate the historic differences in reductionism's substrate

²⁴ "a man who doubts everything material cannot for all that doubt his own existence. From this it follows (il suit) that he, that is his soul, is a being or substance which is not at all corporeal (point du tout corporelle), but whose nature is solely to think (sa nature n'est que de penser), and that this is the first thing, one can know with certainty" (AT I 353: CSMK ss) — Descartes on a letter written to Jean de Silhon in May 1637, at (Cottingham, 1992, p.242); "The intellectual simple natures comprise cognition, doubt, ignorance, the action of the will, and so on [cognitio, dubium, ignorantia, voluntatis actio, AT X 419: CSM I 44]" (Cottingham, 1992, p.126);

²⁵ Alluding to the difference in *Explanatory* or *Constitutive* as defined, explained and cited above.

²⁶ The Epistemological Criticism stems from the idea that — even if ultimate explanations of phenomena do exist at the level of a minimal common substrate — they are inherently inaccessible to human cognition due to our finite mental capacities. As Thomas Nagel (1998) put it in his critique, "If there were a detailed causal explanation at the subatomic level of the movement of a pianist's fingers in the course of playing a Beethoven sonata, we would not be able to hold it in our minds". While this hints at the idea of levels — that I'll soon approach in attempting to solve the Ontological Criticism — taking Nagel's judgement at face value strikes me as both misplaced at solely targeting reductionism and societally condescending. First, it is possible that, given our scale in longevity and size, we might be in the midst of such phenomena we do not recognize, and our same mental restraints would also hinder us at understanding them at a holistic (non-reductionist) level (imagine a music of the cosmos played intergalactically), which make the critique not one of reductionism but one of generalised human accessibility. Then, although "I" surely wouldn't (be able to hold it [the music] in my mind), it is completely plausible that "we" eventually could, as illustrated by the myriad of ever growing scientific accomplishments extrapolating atomic and even lower level information into renders of scales we comprehend. Atop these two points, it is arguable that, from a scientific endeavor standpoint, just because something isn't graspable from an individual perspective, it is not valuable and worth pursuing nevertheless (think of Niels Bohr famous quote "Anyone who is not shocked by quantum theory has not understood it." — in the introduction of "The Historical Development of Quantum Theory, Vol. 3" by Jagdish Mehra and Helmut Reichenberg)

²⁷ Once we have a concept of *domestication* nested at the correct level so as to include its defining characteristic, I am not arguing that a *participle physics* level analysis should not be done on the phenomena: I believe in quite the opposite. However, the utility derives from a higher level understanding of the phenomena

dependency²⁸, utility²⁹ and intentionality³⁰ — which characterises the phenomena — can one speak of domestication? At what level is it ontologically manifested? This is further complicated when speaking about emergent phenomena³¹, such as the positive feedback-loop introduced by *domestication*: increasing agricultural intensification, feed(ing)back into larger human populations, in turn allowing and accelerating further agricultural innovations, markets (can *this* emergent property be described by particle physics?) and technologies that further enhance *domestication*³². Not even mentioning the ontological problems of framing particle physics as a reductionist substrate to phenomenological states³³.

The Downward Causation Criticism refers to the influence of *the whole* on its constituent parts. That is, higher-level phenomena exerting causal effects on lower levels. This is disruptive to, for example, particle physics as a reductionist substrate, because effects would be derived from something other than the reductionist substrate from which causality or explainability ought to emerge³⁴. Thus the underlying assumption of this critique is that higher-level (such as biological, psychological or societal) phenomena can influence lower-level processes without violating physical laws³⁵. Donald Campbell (1974) showcased this within higher-level biological roles, providing the example of soldier ants and how societal functions dictate the structural traits of lower-level entities, like their jaws³⁶. In this instance, although their jaws (lower level than the ant society, but higher level than the reductionist substrate) are still determined by their lower-level properties (their particle physics makeup), their configuration is nevertheless not explained by particle physics alone: to fully account for their configuration one would have to include such higher level strata as the ant colony's *needs*, demographics, and societal *protocols* (Li & Zhou, 2014, p. 300).

Thus, the effort in finding a minimal common denominator substrate between *Cognitive* and *Political* science, from which to derive my analysis, is infused with two seeming limitations of seeking a reductionist approach — first, accounting for non-reducible emergent properties; second, of not limiting explanatory power by framing a substrate too static to account for bilateral causality — plus the general challenge of being compatible with the analysis of two disciplines seemingly occupying two different epistemological realms. This is the challenge setting up the first chapter of this thesis, and from which an analytical lens is derived.

²⁸ "The appreciable metabolic and morphological changes that often accompany behavioral adaptation to the human environment usually lead to a significant dependence on humans for food and shelter" (Driscoll et al., 2009, p. 9974).

²⁹ "Those wolves less afraid of humans scavenged nomadic hunting camps and over time developed utility, initially as guards warning of approaching animals or other nomadic bands and soon thereafter as hunters" (Driscoll et al., 2009, p. 9971).

³⁰ "Artificial selection is a conscious, if unintentional, process, and therefore is generally considered to be effected only by humans" (Driscoll et al., 2009, p. 9972).

³¹ "Emergence might rely on a supposition of indeterminism (...) while the existence and behaviour of such higher-order systems is consistent with the basic laws of physics (...) they are not causally determined by those laws." (Nagel, 1998, p. 6)

³² "Domestication of animal species produced a cycle of positive feedback, which '... sparked ongoing intensification of agricultural production.'" (Green, 2023, p. 436); "Domesticated animals became integral to economies by contributing to transportation, fieldwork, and trade networks." (Green, 2023, p. 436).

³³ If I were to mention it, I would quote: "The color and other features of the experienced table do not really exist in the world independently of us at all: they result merely from the interaction of our sensory apparatus with some physical reality" (Jones, 2013, p. 10).

³⁴ "Downward causation would violate the causal completeness of the subatomic realm, and if biological or mental causes affect the physical, the causal completeness of the physical level." (Jones, 2014, p. 75).

³⁵ "Each level operates with its own independent causes, and the higher-level phenomena can exert a causal influence on the stable processes occurring in their component parts." (Jones, 2014, p. 75); "Downward causation (...) is the claim that in a complex system exhibiting emergence, the higher or macro-level as the locus of emergent phenomena exerts some kind of causal influence downwards on the lower level substrates" (Campbell, 1974, p. 141)

³⁶ "The jaws are so specialized in their function of piercing enemy termites, antler pincers, that a soldier termite cannot feed herself but requires to be fed by other worker termite. But this implied that it was the hierarchical level of a society which had to enter the explanatory principle since selection is now following from such societal factors as the division of labor in the termites' social milieu. (Campbell, 1974, p. 143); that is, "This requires a 'downward' explanation in terms of the ant's role in its society; it cannot be simply explained 'upwardly' by means of its chemical makeup." (Jones, 2014, p. 75)

Chapter 1

Information

1.1 (Information as) Decreases in alphabets of states

To both the Downward Causation and Ontological Criticism, General Systems Theory (GST) offers a trail towards resolution³⁷ by emphasising the structural organization of reality in terms of interconnected and hierarchical nested systems, as opposed to focusing on isolated parts³⁸. *Systems* are here understood as interacting parts that form a coherent entity³⁹. This interaction is central to the organisation and behaviour of the system, distinguishing it from a mere collection of independent parts, and is defined by a seemingly contradictory characteristic: variability within invariance, that is, overall invariance despite the fluctuating and erratic behaviors of its individual components⁴⁰. Weiss (as cited in Smythies & Koestler, 1969, p. 13) expressed this mathematically by stating that “[a given] *complex is a system if the variance of the features of the whole collective is significantly less than the sum of variances of its constituents*”.

As a simple illustration, consider an (attempted) three-dice system versus a cellular system (Mitochondria, Ribosomes, Lysosomes...). When rolling three independent dice, each of the 216 possible outcome combinations occurs with equal probability, and there is no relationship between any one die and any other, so knowing the face of one die says nothing about the faces of the rest. In contrast, cellular organelles exhibit functional dependencies⁴¹ — such as ribosomes requiring mitochondrial ATP for protein synthesis⁴² — that constrain the system to occupy only a subset of the theoretically possible joint configurations. Knowing the energetic state of the mitochondrion narrows what the ribosome can plausibly be doing, while knowing the calcium concentration narrows what the membrane potential can be. Making the cell, unlike the dice, mutually informative across its parts.

Weiss reached for variance to express this intuition because variance was the second-order statistical tool natively available to mid-twentieth-century systems thinking. His substantive insight set me off in the right direction, in that the parts of a system render each other less free than they would be in isolation. Variance, however, turns out to be a *vehicle* that fails on Weiss's own intuition in two distinct ways, and pausing on both is worth doing so because together they explain further limitations to consider while searching for a reductionist substrate.

The first failure is one of *direction*. Variance, as a statistical quantity, measures how spread out a variable's values are around their average. When two variables are summed, the variance of the sum decomposes into the sum of their individual variances plus a

³⁷ Though my search towards this resolution was far and wide, I've curiously encountered it at *home*, since Ludwig von Bertalanffy — an early system's thinking theorist — is also a University of Vienna alumni (one of his happy affiliations, among other deeply unhappy and disappointing ones)

³⁸ “Reality, in the modern conception, appears as a tremendous hierarchical order of organized entities, leading, in a superposition of many levels, from physical and chemical to biological and sociological systems. Unity of Science is granted, not by a utopian reduction of all sciences to physics and chemistry, but by the structural uniformities of the different levels of reality.” (Von Bertalanffy, 1969, p. 88)

³⁹ “A system or ‘organized complexity’ may be circumscribed by the existence of ‘strong interactions’ (...) or interactions which are ‘nontrivial.’” (von Bertalanffy, 1969, p. 55).

⁴⁰ “The basic characteristic of a system is its essential invariance beyond the much more variant flux and fluctuations of its elements or constituents” (Weiss, as cited in Smythies & Koestler, 1969, p. 13).

⁴¹ (Lindqvist et al., 2018); (Brown, 1992); (Rangaraju et al., 2019)

⁴² (Rangaraju et al., 2019); (Rossoll & Bassell, 2019)

correction term that captures how much they tend to move together. That correction term is positive when the variables co-vary in the same direction (one rises as the other rises) and negative when they *trade off* (one rises as the other falls). For the variance of the sum to fall below the sum of individual variances — which is what Weiss's criteria identifies as *systemhood* — the correction term has to be negative, meaning the components compensate each other. Weiss's criterion, taken literally, therefore picks out as systems those collections whose components are *anti-correlated*. But the systems this thesis care about often exhibit the opposite sign. For example, cellular organelles tend to be positively coupled, ribosomes are most active when mitochondria are producing the most ATP, calcium signalling and membrane potential covary in the same direction during cellular activation... Under Weiss's literal criterion, a cell undergoing coordinated activation would have higher joint variability than the sum of its parts and would not register as a system at all. This is a diagnosis Weiss himself would not have wanted⁴³, and an artefact of using a second-order signed quantity to capture an intuition about mutual constraint.

The second failure is one of *scope*. Variance only registers the kind of dependence where the relationship between variables is roughly linear, where one variable goes up by some amount and the other goes up or down by some amount in proportion. Two components that are tightly coupled in non-linear ways (for instance, a variable that is symmetrically distributed around zero and that variable's square) can have zero linear correlation and yet one fully determines the other. Biological, cognitive, and governance systems can be expected to be composed, at least in part, of such non-linear couplings.

While I highly empathise with the direction of Weiss' substantive insight — being almost axiomatic *a priori* for me that components of a system are mutually less *free*⁴⁴ than they would be in isolation — the formal *vehicle* of this intuition needs to be one that, first, registers all forms of mutual constraint regardless of the sign of any particular pairwise relationship, and second, registers non-linear and higher-order dependencies that variance is structurally blind to.

1.1.1 When I learned about probability spaces

Most of the readings I had encountered that used probabilistic language — mainly on free-energy minimisation, active inference and predictive coding — would write expressions like “ $p(x)$ or “ $P(X|Y)$ ”. Until this thesis, I had proceeded without ever explicitly committing to understanding what kind of object p or P were supposed to be representing. For a reader from a Law background graduating in CogSci⁴⁵, I understood probabilistic statements functionally yet with a low-key confusion that every probabilistic expression seemed to float free from a concrete referent I could solidly grasp. So it was unclear which results depended on which assumptions about the underlying objects. Only upon seeking a deep understanding of Friston's measure-theoretic framing of Markov blankets did I eventually grok it, through Cover and Thomas's (2006) information-theoretic textbook. A probability space is a triple of three pieces: the sample space, the collection of events, and the probability measure.

⁴³ According to the spirit of his GST intuition

⁴⁴ In terms of degrees of freedom

⁴⁵ Being my background

The sample space

The first piece is the set of all *microstates* available to a system under analysis. If what one is analysing is a die, the sample space is the set of faces {1, 2, 3, 4, 5, 6}. If one is analysing a cell, the sample space is something like the set of all possible molecular configurations the cell could be in, that is, while astronomically larger conceptually the same exhaustive list of "ways this system could be in". While analysing a governance system, the sample space might be the set of all possible configurations of laws, citizens, and institutional states (and even, as this thesis will further develop, their nested states). The crucial point is that the sample space is fixed by the inquiry, being functional and relational to what the analysis is about, so there is no requirement to commit to any *ultimate base level*, only to the level the inquiry requires.

The collection of events

The second piece is the collection of subsets of the sample space to which a probability can be coherently assigned. This need to be specified separately because for some sample spaces, like continuous ones, one cannot consistently assign probabilities to every possible subset⁴⁶, though for finite sample spaces this collection can simply be all subsets of the sample space⁴⁷. Functionally, one can think of this collection as "the events one can ask about", questions of the form "is the system in some particular subset of states?" that admit well-defined answers. When a system "learns"⁴⁸ or acquires knowledge, then uncertainty is reduced corresponding to a distribution being concentrated by the arrival of new data (an observation, a measurement, the completion of a process).

The probability measure

The third piece is the function that assigns a number between zero and one to each event, capturing how likely that event is⁴⁹.

These three pieces work together as a unit. The sample space tells us what the universe of microstates the inquiry admits is, while the collection of events functionally tells us what we are allowed to ask about and the probability measure tells us how likely each of those questions is to be answered yes. Together, it expresses a universe of possibilities framed through a collection of well-formed questions each with a weighting of probability. This matters because a "likelihood of a microstate" or "uncertainty about which configuration the system is in" is how one can understand the constraints that compress the alphabet of states that make subsystems *less free* (as per Weiss').

1.1.2 "Information" Desambiguation

"As it turns out, flipping a coin offers just about the simplest possible picture of what information is all about. A coin can turn up either heads or

⁴⁶ A result from measure theory. The technical issue being that some subsets are non-measurable, leading to paradoxes if one tries to assign probabilities to them

⁴⁷ For continuous sample spaces, it has to be specified, though I will not pursue that level of detail here

⁴⁸ Funnily, I also came to the conclusion, while *leaning* this, that when a system "learns" or "acquires knowledge", it moves from uncertainty about which state it should be (behaviour and action comprehend constellations of microstates compatible with said physical and mental operations) to greater certainty. Before learning, the system might have a broad, relatively uniform probability distribution across many possible response states. After learning, this distribution becomes concentrated around fewer, more appropriate states.

⁴⁹ This measure has to satisfy a few consistency requirements: the probability of the whole sample space is one; the probabilities of disjoint events add up; and the measure is well-behaved under countable unions

tails. Two possibilities, equally likely. When you catch the coin and remove the covering hand, you find out which of the two possibilities it is. (...) you have acquired a single bit. A bit doesn't have to involve coins. A bit can be represented by a lightbulb—on or off. By an arrow, pointing up or down. By a ball, spinning clockwise or counterclockwise. Any choice from two equally likely possibilities is a bit."

(Siegfried, 2000, p.2)

Bertalanffy's understanding of information is rooted in Claude Shannon's *mathematical theory of information*, in which *information* is defined to quantify the reduction of uncertainty within a system (Shannon, 1948). Yet Cover and Thomas (2006) keep them apart cleanly three ways in which the word *Information* is used across to express information-theoretic quantities / roles.

The first concept is negentropy, as the constraint content of a single distribution. Take a level of description, fix the probability space at that level and compare the system's actual distribution over microstates⁵⁰ to the maximum-entropy reference (of a uniform distribution), then negentropy measures how far the actual distribution departs from that uniform, *flat*, reference distribution. For example, a fair die has zero negentropy. Negentropy is then what constitutes an alphabet compression, the fewer effectively-occupied states a single distribution concentrates on, the higher its negentropy. A weighted die, that suppresses other faces in favor of a specific one, would have progressively more negentropy the more weighted that face is. It is important to understand negentropy as a property of distribution. This is the closer information definition to Bertalanffy's understanding of Weiss systems as expressing a departure from random distribution, expressing this directly when he writes that "*entropy, as we have already heard, is a measure of disorder; hence negative entropy or information is a measure of order or of organization since the latter, compared to distribution at random, is an improbable state*" (von Bertalanffy, 1969, p. 42).

The second is mutual information, as the coupling between two random variables. Take two random variables — for instance, one summarising the state of a higher level and one summarising the state of a lower level — asking "how much does knowing the value of one reduce uncertainty about the value of the other?" yield information. This is the right quantity for talking about how much one part of a system informs another part and for talking about how much higher-level constraints inform lower-level configurations (a slight derivation for the informational setting above, given that negentropy on one level might inform outcomes on another level⁵¹: such as weighted die and likelihood of set winnings in casino games). So, while not the same as negentropy, mutual information across levels is related.

The third, being the coupling among the components of a system at one level (Watanabe, 1960), is total correlation. If a system at a given level decomposes into components — such a cell into organelles or country into citizens — total correlation measures how far the joint distribution of all components departs from the product of their individual marginal distributions, being zero exactly when all components are jointly independent.

Total correlation also generalises Weiss's intuition, in that when the parts of a system constrain each other, knowing one part tells you something about the others, and the joint distribution concentrates onto a subset of the full Cartesian product of component states,

⁵⁰ 1.1.1 - The sample space

⁵¹ More on levels on the follow subsections

with total correlation registering whether the components are positively or negatively correlated or whether their dependence is linear or non-linear.

These complementary definitions of information, in the manners illustrated above, carry the substantive GST claim, and although individually unpacked above, *information* will be used throughout this thesis to signify reduction in uncertainty across any or each of these framings accordingly.

1.1.3 System, revisited & defined

Considering the above, Weiss's intuition can be restated with further detail. A multi-component level constitutes a system, to this thesis, when its total correlation is positive, in the sense that the joint distribution of its components is concentrated on a strictly smaller alphabet than the Cartesian product of the component alphabets would allow. Whereas independence implies zero total correlation, dependence among components, whether positive or negative, linear or non-linear, drives total correlation above zero, because total correlation detects all forms of statistical dependence, while variance super-additivity only detects negative net covariance. The three-dice collection has zero total correlation and is therefore is not a system, whereas the cell has positive total correlation across its organelles and is.

Thus, in a system (for this revised definition) the correlated interactions between components lead to more concentrated probability distributions over system microstates than would occur under statistical independence. This is in line with GST definition of systems in relation to overall invariance — thus, axiomatically — to order and maintenance, to dampened disorder, and so, lower entropy⁵². That is, given that systems⁵³ — at any given level — transition among spectrums of possible microstates across time, and that these microstates can be probabilistically described, it is possible to ask the question "to which degree am I certain about the microstate in which the system currently resides?".

The more uniform the microstate probabilities across the available alphabet, the higher the entropy, whereas the more concentrated the distribution around specific configurations, the lower the entropy. Following this line of reason, systems are thus informational manifestations across different microstate alphabets *restricted* throughout diverse types of constraint (consider the differences between quantum selection rules constraining electronic configurations when restricted by an atom system versus constraints in mitochondria alphabets of states when part of a cellular system). A microstate being one fully specified, fine-grained configuration, or state, of any given system or its parts. While for a die, a microstate is simply the face showing after a throw, having simply 6 available microstates in its microstate alphabet (i.e, the landscape of all potential microstate configurations), for a mitochondrion, a microstate represents the detailed configurational snapshot of its molecular variables (specific enzyme conformations, local ATP/ADP concentration ratios, calcium ion distributions, membrane potential gradients, etc), and so, while its theoretical alphabet is astronomically large, its functional alphabet (compatible with mitochondrial integrity) is dramatically constrained by the biochemical requirements for

⁵² In standard thermodynamics, as applied to heat engines, say, entropy is a measure of the degree of disorder in a system (Davies & Gregersen, 2010, p.77)

⁵³ The link between thermodynamic order (the electron's accessible volume, the mitochondrion's biochemical configurations) and Shannon entropy (uncertainty over an alphabet of states) is the *standard maximum-entropy bridge* developed by Jaynes (1957) and Landauer (1961)

maintaining overall mitochondrial viability. Thus a system's alphabet of available microstates is dependent upon the degrees of freedom of its parts compatible with system integrity, i.e., up to the point at which a configuration leads to system disintegration (electrons ionizing from atoms, enzymes denaturing in mitochondria...). Thus systems concentrate probability distributions around fewer, more predictable configurations. The viable microstate alphabet contracts from its theoretical maximum.

Though what constitutes a microstate — and thus the probability space — depends critically on the level and scale of description at which the system is situated, the underlying informational process remains constant as the degree to which an attempted system exhibits or not concentrated probability distributions over their accessible microstates, due to constraints that limit random variation.

This reveals information's unique suitability as a cross-level⁵⁴ *reductionist substrate*. System integrity — that is, the viable low level alphabet of microstates compatible with higher level stability — depend on operations' that manifest different constraint types, according to distinct microstate definitions and their corresponding constraint mechanisms (e.g. atomic orbitals versus versus cultural taboos), that diminish state space configurations regardless of substrate. This informational substrate provides as a universal metric the degree to which the components of a system are mutually *less free* than they would be in isolation.

1.1.4 Information across *levels*

Bertalanffy (1969) and the systems-theoretic literature he opened spoke of these as "levels of organisation", however what has been developed so far operates within a single *level* of description, a cell qualifies as a system because its organelles are mutually *less free* than they would be in isolation, and thus the joint distribution over their states concentrates, total correlation is positive and the effective alphabet of accessible configurations is smaller than independence would allow. But the same cell can be described at the level of its molecular configurations or at the level of its metabolic states. The information-theoretic quantities defined in 1.1.2 are all defined relative to a probability space, and a probability space is defined relative to a level and scale of description. Moving from "every possible molecular configuration in integrity" to "respiring vs quiescent vs apoptotic" is moving to a different set of probability spaces with different alphabets of microstates composed by the same subsystems. The claim above⁵⁵ is resting on the informational structure visible at one level bearing a determinate, formally specifiable relationship to the informational structure visible at another.

The constraint structure that makes the cell a system is equally present at both levels but it registers differently within each probability space. Two microstates are equivalent at a higher level when, from that higher level's vantage point, they count as *the same thing*. The function performing the grouping is what mathematicians call a coarse-graining, i.e., a map that takes microstates as input and returns macrostate labels as output. "The cell at the molecular level" is a probability distribution over molecular configurations while "the cell at the metabolic level" is what you get when you apply a function that sends each molecular configuration to a metabolic label. Returning to the dice example, with a base sample space of six elements, with each face having probability one-sixth. To define a coarse-graining that

⁵⁴ More on what constitutes a level on the following sections

⁵⁵ That information serves as a cross-level reductionist substrate

sends each face to one of two labels, "odd" or "even", the higher-level sample space then has two elements. The probability of "odd" at the higher level is just the total base-level probability of one, three, and five, which is three-sixths or one-half (of course, same for "even").

Coarse-graining is the procedure of partitioning a fine-grained microstate alphabet into equivalence classes, where each class groups together all the microstates that are indistinguishable at the coarser level, so that each class then becomes a single *macrostate* at that level. Robertson (2020, p. 564) puts it collapse in alphabets of state elegantly, referring to them as "*empirically indistinguishable*", not because their measurement is limited, but because the distinctions between microstates within a class carry no weight at the higher level, meaning that, once grouped, the probability that the coarse-grained system assigns to any macrostate is just the sum of the fine-grained probabilities over the equivalence class beneath it where a distribution over a larger alphabet is replaced by a distribution over a smaller one. What the coarse-graining operation collapses is therefore detail that is, for the purposes of the higher level, probabilistically *irrelevant* since it is subsumed into the broader class distribution. As Robertson (2020) puts it, the macrostate description is a new variable in relation to a higher description level. Knowing the cell is respiring is not a worse version of knowing its molecular configuration, but rather a different piece of information about a set molecular alphabet of states compatible both with respiration and with cellular integrity.

From a formal perspective, the entropy of the fine-grained description equals the entropy of the coarse-grained description, plus the average within-macrostate entropy, the residual uncertainty about which fine-grained microstate, within its macrostate, the system currently occupies. Total fine-grained uncertainty equals uncertainty about which macrostate the system is in, plus, on average, leftover uncertainty within those macrostates. Wherever the components of a level mutually constrain each other⁵⁶ and that level is coarse-grained into macrostates compatible with the system's constraints, the resulting macrostate alphabet is provably smaller than the microstate alphabet. However, levels do nest. If we apply a second coarse-graining that further aggregates the first-level macrostates into even coarser categories, the composition of the two coarse-grainings gives a coarse-graining all the way from the base to the second level. The second-level probability measure is the pushforward of the base measure through this composition. Each level in the hierarchy is one function-composition further from the base than the level below it. Each level's probability content is determined by the base probability content together with the chain of compositions that connects them. The structure is thus recursive, whenever the framework needs another level above, another coarse-graining suffices and the higher-level probabilistic content follows by pushforward.

⁵⁶ i.e., as per 1.1.2, positive total correlation

1.2 Systems as historical restraints-cascades

The previous section establishes information as the measurable reduction of uncertainty over a system's alphabet of microstates, in which components mutually constrain each other leading their joint distribution to concentrate onto a smaller alphabet. It was also established that through coarse-graining, microstates of lower level systems can be nested towards a description of states corresponding to a higher level system alphabet of states.

What is still missing is a dynamical account of how such constraints accumulate through time to support open-ended complexity. Assembly Theory (AT) (Sharma et al., 2023) provides that missing link by reconceptualising what physically counts as an “object”, not as a point particle but as an entity defined by its (possible) formation histories in an assembly space⁵⁷. The key shift is that an object's identity is no longer exhausted by its present configuration, not considered as frozen in a slice of time. From an AT vantage point, the existence of many identical, multi-step objects signals selection (restraints, and thus information gain) acting on historically realised construction paths.

Understanding information in physics as the resolving of constraints in paths towards Assembly Theory's (reconceptualisation of) objects, transforms information into the primary substrate for understanding structural organisation, revealing that what we traditionally consider discrete “things” — particles, molecules, organisms and so on — are actually stable patterns of historically accumulated constraint propagation.

In this framework, each object carries within its structure the informational content of its *assembly history* as quantifiable reductions in microstate alphabets. This is made clearer by AT's key concepts of *assembly index* and *copy number*. In this framework, *assembly index* is the length of a minimal sequence of operations (the shortest construction path) needed to build an object from previous *objects* as building blocks, functionally operationalising the depth of cascading historical constraints required to reach any given object's assembly. The *copy number* records how many identical instances of that object exist. While high copy numbers of objects with low assembly index are expected, high copy numbers of objects with high assembly indexes require forms of selection. The argument being essentially combinatorial, sustaining that the eventuation of objects with high *assembly indexes* is very hard via combinatorial explosion alone, but once a *path* is discovered, making copies along (or streaming off) that path becomes much more informationally feasible.

This is where Assembly Theory highly converges with the approach taken above, by separating “assembly possible” spaces — i.e., “*the space of physically possible objects, which can be generated by means of the combinatorial expansion of all the known physical rules of object construction and allowing all rules to be available at every step to every object*” (Sharma et al., 2023, p. 325) — from “*assembly contingent*” ones which describe “*the possible space of objects where history, and selection on that history, matter*”, that is, the historically realised, path-dependent subspace, where bases towards higher assembly indexes are made possible. While the assembly index measures the depth of constraint cascades required for object assembly, copy number indicates the success of those constraints in propagating stable patterns. The assembly route towards an object — its

⁵⁷ “In AT, objects are not considered as point particles (as in most physics), but are defined by the histories of their formation as an intrinsic property, mapped as an assembly space. The assembly space is defined as the pathway by which a given object can be built from elementary building blocks, using only recursive operations. For the shortest path, the assembly space captures the minimal memory, in terms of the minimal number of operations necessary to construct an observed object based on objects that could have existed in its past.” (Sharma et al., 2023, p. 321)

history — functions as a memory, perspectivising any given system as a historical restraints-cascade when its components are organised so that (earlier established) constraints shrink the accessible microstate alphabet of other components (Weiss property), and the system retains those constraints as a *memory* that enables channeling subsequent exploration on assembly space, enabling the realisation of higher-assembly objects without having to exhaust all lower-level combinatorics, that is, offering a route towards higher assemblies, higher information gain.

Binding together the section above (entropy-reducing constraints over alphabets of states) with this historical picture (path-dependent construction in assembly spaces) offered by AT, provides a useful framework from which to derive this thesis' analysis. A constraint, in the sense relevant here, being anything that reduces the number of states a system can occupy, from chemical bonding rules, physical conservation laws, biological viability requirements, socio-political regulations... each of these eliminates some configurations of the underlying components, or makes some configurations extremely unlikely. This allows one to think of constraints as performing a kind of compression on the original alphabet of states. Where originally the joint distribution over components ran over the full Cartesian product of component alphabets, after constraints are applied, the joint distribution runs over a smaller alphabet and the probability mass redistributes, concentrating on the configurations that satisfy the constraints, while vanishing on those that don't. The total correlation among the components rises, the entropy of the joint distribution falls, and as a result entropy decreases.

Once these less entropic states are achieved, the system state is available for reuse along its assembly path. AT conceptualises this as systemic memory, as the system conserving sets of constraints that narrows future construction choices while stabilising assembly solutions to viable objects. So an assembly path, once discovered and stabilised, functions as a selectional bias on subsequent construction, with configurations that would have been combinatorially astronomical to reach from scratch becoming routinely accessible because the path that reaches them has been crystallised. In this manner, one can understand routes in constraints as selection, as assembly paths are reused high assembly indexes can be achieved without having to emerge *ad ovo*. It could be more intuitive to grasp this concept using its counterfactual: without mutual restrictions, parts remain independent, and so the joint alphabets of sub-systems factorise and the system does not gain predictability (no entropy reduction and no system in the Weiss/total-correlation sense), without memory, even strong and stable constraints can't repeatedly reuse successful assemblies, so assemblies remain undirected and shallow on *assembly index*.

This creates a framework where the fundamental question shifts from "*what are things made of?*" to "*what informational constraints were necessary to construct and maintain this configuration?*", making constraint propagation and alphabet compression the foundational organisational principles underlying structural richness from atomic assemblies to biological systems and beyond.

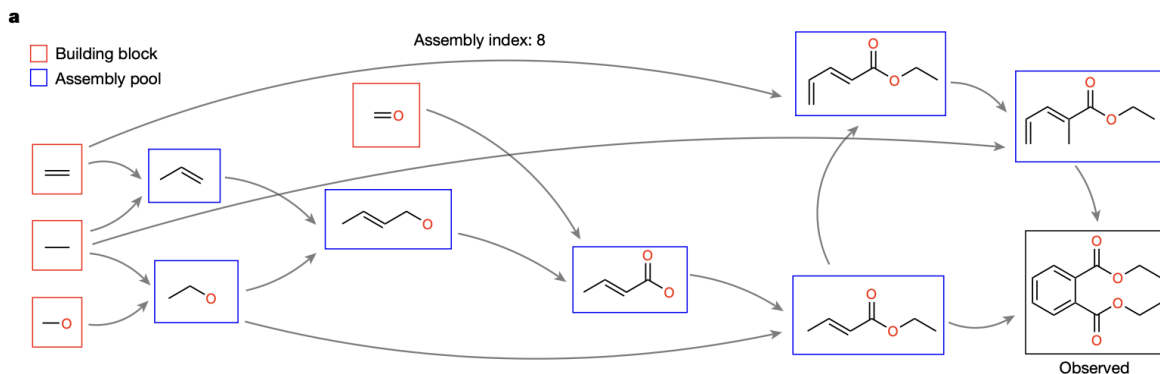


Figure 1: Illustration of Assembly Index (Sharma et al., 2023, p. 322). As per the figure's original legend: "AT is generalizable to different classes of objects, illustrated here for three different general types. a, Assembly pathway to construct diethyl phthalate molecule considering molecular bonds as the building blocks. The figure shows the pathway starting with the irreducible constructs to create the molecule with assembly index 8."

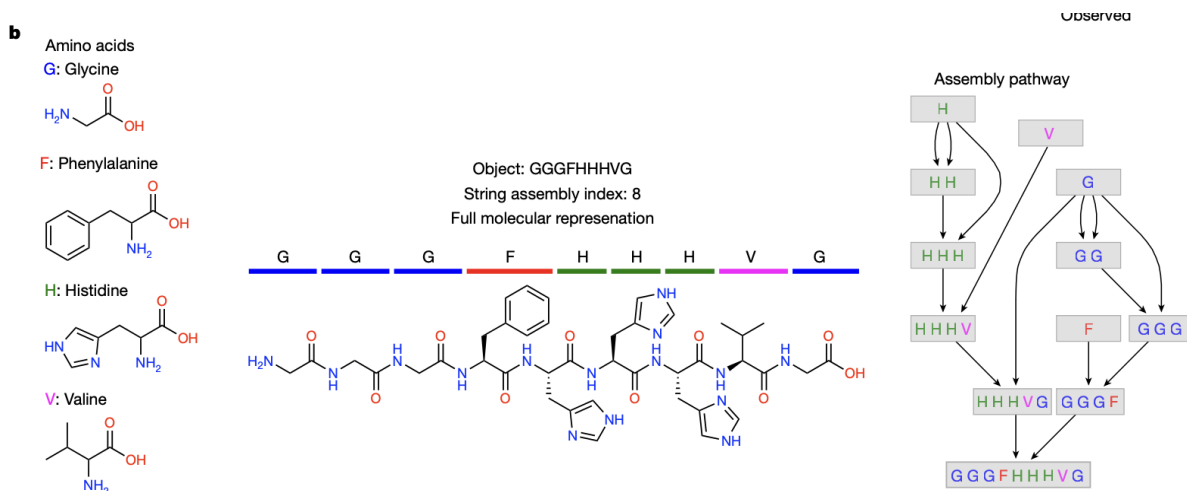


Figure 2: Illustration of Assembly Index (Sharma et al., 2023, p. 322). As per the figure's original legend: "Assembly pathway of a peptide chain by considering building blocks as strings. Left, four amino acids as building blocks. Middle, the actual object and its representation as a string. Right, assembly pathway to construct the string. c, Generalized assembly pathway of an object comprising discrete components."

1.2.1 Shared sub-paths and the role of assembly bottlenecks

Distinct objects often share subpaths, when different targets reuse the same subassemblies. Achieving system stability or solving an integrity problem often provides the *memory* to many future iterations. The peptide bond of Figure 2 showcase how amino acids could in principle arrange themselves in astronomically numerous configurations, but the peptide bond is a convergent constraint-satisfaction solution that, once discovered, becomes a reusable subassembly that countless protein structures incorporate, dramatically reducing the search space for viable molecular architectures.

To this dependency on previous assembly solutions from which other future assemblies with higher *assembly index* are derived this thesis refers to as *assembly*

bottleneck(s). A substructure, within the assembly space, is an *assembly bottleneck* of degree k in the sense that its indices of at least k distinct objects are realised only by paths that pass through that substructure. Equivalently, an assembly bottleneck is a vertex of high betweenness in the directed acyclic graph of minimum-path realisations, restricted to those paths. Three properties follow from this definition. First, an assembly bottleneck exhibits convergent constraint satisfaction, that is, multiple independent constraint sets find common resolution at it (this is what k large means). Second, it exhibits reusability amplification, meaning that once established, the bottleneck can be incorporated into numerous higher-order assemblies without requiring re-derivation of its constituent constraints (this is what its position high in the path-graph means). Third, it exhibits informational compression, in which the vast microstate alphabet of possible arrangements at and below the bottleneck contracts to a functionally viable subset on the post-bottleneck side (this is the §1.1.4 quantity, applied locally). Co-assembly and joint assembly via such Assembly Bottlenecks is why one can think of assembly objects as modular, allowing systems to advance along each other's assembly pathway, as each successful assembly path holds the potential to be a reusable step towards other assemblies. This is why systems that store and reuse constraints can reach high *index assemblies* without first exhausting all the shallow possibilities, without a system *needing* to generate complexity ad ovo *it can* instead stack and reapply reductions of uncertainty so that each step channels the next.

1.2.2 Assembly across scales and substrates

Most importantly, deploying General Systems Theory alongside Assembly Theory allows treating objects at their informational level, not as the discrete, temporally bounded entities — i.e. existing at a given particular moment which freezes reality to a biased, discrete human-based experience of time — towards a flux of stable patterns of accumulated constraints that have successfully propagated through assembly space. This has several implications which please my anthropocentric-averse epistemology. First, it dissolves the traditional subject-object distinction that treats systems as collections of independent parts, to instead regard observable configuration as cascading historical constraints where earlier selections channel later possibilities. This is fundamental, as I personally suspect it is intuitively hard for humans to regard post-human assemblies (such as Governance Assemblies, as this thesis argues) as belonging to an informational flux of constraint propagation that eliminated incompatible configurations from its microstate alphabet. Second, it explains how complexity emerges from leveraging previously established constraint patterns (the assembly bottlenecks defined above) to access higher assembly indices. The apparent violation of entropy increase in unstable post-bottleneck assemblies is resolved by recognising that no system is closed, but instead they inherit informational constraints from their assembly histories, allowing them to occupy low-entropy configurations (that would be combinatorially impossible to reach ad ovo) that allow experimentation with novel assemblies which might be of higher entropy up to the point of novel found stability (a new assembly bottleneck). Third, the modular reusability inherent in assembly paths explains why assemblies and integrity solutions often have striking convergent patterns across different domains.

Chlorophyll, for example, is a bottleneck dependency for all *plant* assemblies⁵⁸, yet *itself as an entity* — the minimum number of recursive, compositional joining operations required to assemble chlorophyll — is dependent on the pathways of previous structures taken to assemble it, which are themselves constrained by the components that already existed and have been formed through prior assembly steps. Chlorophyll is way too complex and sophisticated to ever be spontaneously and arbitrarily materialised in random particle interactions, instead, it results from a specific trajectory. In the absolute vastness of *potential combinatorial space*, considering *all* that can be conceived physically, if the random formation of chlorophyll is astronomically improbable, then the one for trees is even more unimaginable. Thus assembly threads this enormous *combinatorial space* through sets of constraint pathways pruning state-space in two manners: history dependency at the lower level and fitness / resilience / integrity at the higher level⁵⁹. This second pruning direction — the higher-level fitness criterion shaping which lower-level paths get reused — is the topic of section 1.3.

From this perspective, any process or set of constraints that leads to the enrichment of assemblies by promoting their arrangement towards higher indexes, can be regarded as *selective* — even in non-living matter (such as in minerals or virus) — considering *fit* the variables *integrity* (degree to which the arrangement retains its structural coherence) and *copy number* (abundance, or number of occurrences, of identical assemblies). *Life* thus represents a specially highly effective, historically contingent process of discovering and propagating specific, functional, high assembly index arrangements, though Darwinian evolution would be *a*, instead of *the*, selector of assemblies (with life being a bottleneck solutions towards assembly section and integrity in post equilibrium composites).

As hinted at above, I believe the dispassionate systemic analysis of governance solutions is hindered by the traditional approaches to understanding objects and systems, which remain trapped within human temporal experience and physical scale. For me, this is critical because a more atemporal perspective helps in extending the dynamics showcased above beyond physics and biology into social institutions, technological systems, and cultural patterns, all representing constraint-cascades that maintain their organization by continuously channeling possibilities toward compatible configurations.

In this framework, memory and remembering is not teleological thus not purpose-driven, but rather a stored property of the object and its historicity itself, encoded by the pathway that generated it, that manifests the object's prevalence. And systems of high informational content are stable trajectories through assembly space instead of objects in the static conventional sense. Whether analysing atomic bonding, cellular metabolism, neural processing, social organization, or political institutions, the same principles of systems emerging and maintain themselves by propagating constraint patterns that reduce uncertainty in their operational domains, while providing assembly bottlenecks that enable future complexity, is maintained. Systems are then understood as information-processing

⁵⁸ "Given that photon-electron conversion efficiency (PECE) is not the same as photosynthetic efficiency (PE) after the comparison of the final carbohydrate energy to the absorbed photon energy, it is now regarded that the light-driven EET is a bottleneck of PETC (Croce and van Amerongen, 2020). To this end, Chl homeostasis and Chl-mediated EET and CS are critical to the survival of oxygenic photosynthetic organisms" (Gao et al., 2024)

⁵⁹ If I may take a (quite big) tangent to extrapolate immensely into a realm outside the explored and cited on the chapters above and beyond the scope of this thesis, it is my intuition that physical laws and properties we know and assume eternalistic today, might arrive through similar mechanics, in the sense that it is possible that other universes with other physical laws exist, however a portion of these end up with assemblies of physical laws which would be nonviable, that is, our physical laws themselves might have dependency routes too, and like there exists a phantom repository or molecular assemblies that never consolidated themselves as viable solutions and thus are extinct from today's routes in assembly, there can be such phantom repository for universes whose assemblies of physical laws were never viable enough to sustain gravity, magnetism, quarks, planets, computation, trees, penguins...

patterns that have achieved stable self-propagation through assembly space exploration, as crystallisations of successful constraint cascades.

1.3 Singularity-nodes as sites of level-formation

Having illustrated how assembly is path-dependent, cumulative and how earlier-stabilised constraints serve as memory enabling higher-index constructions, what I haven't specified yet is the principle by which some paths get crystallised into reusable memory while the overwhelming majority do not. The combinatorial space of physically possible objects is astronomical⁶⁰, while the "assembly contingent" space — the historically realised one — is by comparison vanishingly thin. Sharma et al. (2023, p. 325) draw the distinction between "*assembly possible*" and "*assembly contingent*"⁶¹. This can be operationalised by the sections above. First, the components of a system at any given level must mutually constrain each other for that level to register as a system at all⁶². Second, the constraint configuration that produces such mutual constraint must have arisen through some path in assembly space⁶³. Third, the path becomes available as memory — that is, as a reusable bottleneck for subsequent assemblies — only if the configuration at its terminal node both persists long enough, and reemerges with sufficient fidelity, to participate in further assembly events.

1.3.1 Persistence (without teleology)

The vocabulary adopted, such as "*fitness*" and "*selection*", carries a teleological residue inherited from the way evolutionary biology gets popularised, with a public representation of traits being "*selected for*" because they confer some advantages or gains, and so the language pulls toward notions of intentionality or even of agency where none ought to be implied. Assembly Theory can be understood as radically distant from these notions, in a way that will confuse laymen with a more teleological cultural understanding of selection, as having it be a property of the assembly graph. What gets *selected* is the path that gets reused, regardless of why.

For this thesis this neutralisation matters especially since the framework reaches into systems where talk of "fitness" risks importing the very anthropomorphism section 1.1 set out to avoid⁶⁴. I therefore mean by *selection* something narrowly statistical, the reweighting of probability mass over assembly paths in favour of those whose terminal configurations persist and propagate. Where *persistence* is the maintenance of structural coherence under the perturbations the configuration actually encounters, and *propagation* is copy number, that is, the number of distinct realisations of the same configuration that subsequently arise.

⁶⁰ On the chemical universe of nucleic acids specifically, Walker notes that "*the chemical universe of nucleic acid structures is immense*" yet "*constraints can nonetheless be imposed on which nucleic acids are viable precursors to RNA as only certain combinations of nucleic acids will mediate information transfer.*" (Walker, 2017).

⁶¹ As per above, the assembly contingent space is "*the possible space of objects where history, and selection on that history, matter*".

⁶² Positive total correlation, 1.1.3

⁶³ See section 1.2

⁶⁴ Importing teleology at the bottom of the framework would make it impossible to extract the framework cleanly at the top. Section 1.1's reductionist substrate is meant to do work all the way up the assembly hierarchy, including at levels where teleological vocabulary becomes misleading.

Defining these two variables (integrity and copy number) jointly what counts as *fitness* in this framework. Neither requires that anything intend the configuration to persist or that any agent assess its viability. A crystal lattice that resists thermal perturbation has integrity in this sense, and so does a peptide bond that survives within its biochemical milieu, while a regulatory norm that survives the institutional cycles within which it operates shares the same property. What unites these across substrates is that, once formed, they continue to *obtain*, and once they continue to *obtain*, the path that produced them becomes available as substrate for subsequent assembly.

Any process that reweights the probability mass over assembly paths in favour of those configurations which persist and propagate would constitute *selectors*. Of course, Darwinian evolution is a particularly potent and sophisticated *selector mechanism* in the biological milieu (with processes non-existent up to this assembly index) but the underlying principle is more general than this single high assembly index instantiation.

1.3.2 Formalising *viability*

Two principled formalisations are available for *persistence* and *propagation* to identify the same phenomenon at different levels of abstraction. They both replace the free thresholds that intuitive readings of system-formation often introduce — the moment at which "enough" constraint has accumulated, or "enough" mutual information has built up — with intrinsic features of the joint distribution itself. Working through them requires picking up a few pieces of theory.

Phase transitions in constrained ensembles

The first formalisation comes from statistical mechanics, where the question of which configurations a constrained system tends to occupy is answered through a single quantity called the *free energy*⁶⁵ (Goldenfeld, 1992; Sethna, 2006). For our purposes, free energy can be understood functionally, without committing to its full thermodynamic apparatus. It is a number, attached to a constrained ensemble, that summarises the trade-off the system makes between two competing pressures, the pressure to minimise its energy (which favours ordered configurations, the cultural representation of the universe being *economical* or *frugal* is helpful in grokking this concept) and the pressure to maximise its entropy (which favours disordered ones, the individual experience of making a mess being easier than cleaning up a mess also facilitates understanding). When external conditions favour order, the free energy is minimised by ordered macrostates, while when they favour disorder, it is minimised by disordered ones. The configurations a constrained system will actually be found in, with high probability, are those that minimise its free energy under the prevailing conditions. The free energy is therefore the natural object for asking which macrostates a system "*prefers*" under a given constraint structure, where *prefers* means nothing more than concentrates probability mass on.

What matters for selection is how the free energy depends on the parameters that govern the constraints, be this temperature, pressure, an external field, a coupling strength... The set of distributions one obtains by sweeping such a parameter across its range, with everything else held fixed, is what I will call the *constraint family*, a one-parameter family of distributions over the same components, indexed by the parameter being swept. As any

⁶⁵ Further developed when presenting Active Inference in the next section

such parameter changes, the free energy of the constraint family changes too. For most parameter values the dependence is smooth, with small changes in the parameter producing small, continuous changes in the free energy and in the system's preferred macrostate. At certain critical parameter values, however, the dependence becomes non-analytic, meaning, the free energy (viewed as a function of the parameter) fails to have a smooth power-series expansion at that point, with derivatives of some order becoming discontinuous or diverging (Goldenfeld, 1992; Sethna, 2006). What this means concretely, beyond the formal definition, is that small changes in the parameter on either side of a non-analytic point produce qualitatively new macrostate structure on the other side, with new macrostates becoming stable, previously stable ones become inaccessible, and the structure of the configuration space the system effectively explores reorganising. These critical parameter values are *phase transitions*, and they are the standard formal concept (with a century of development behind it) for thinking about qualitative change in the structure of constrained systems.

The location of a phase transition is fixed by the structure of the constraints themselves and the geometry of the configuration space. This is what gives the present framework a substrate-level criteria for when a configuration counts as having achieved *persistence*. A configuration that sits at a stable phase, one whose constraint family places it at a smooth, low-free-energy region of the landscape and not at, or near, a phase transition, resists perturbations in the parameters that govern its surroundings. Such configurations propagate because they are robust, and they are robust because they sit at minima of the free-energy functional far from its slopes or at the points where it ceases to be smooth.

Information bottleneck bifurcations

The second formalisation does structurally the same kind of work, but without requiring physical free energy. Tishby, Pereira, and Bialek (1999) developed it as a parameter-free way to identify the natural levels of granularity in any joint distribution. Its apparatus is the *Lagrangian*, a standard tool from constrained optimisation in which two (or more) competing objectives are combined into a single function whose minimisation simultaneously addresses both, by weighting one against the other through a control parameter (Boyd & Vandenberghe, 2004).

Where the free energy combines energy minimisation with entropy maximisation under physical constraints, the information-bottleneck Lagrangian combines two purely informational objectives that pull in opposite directions. On one side it asks for compression, starting with a fine-grained variable X (the high-resolution description one has in hand), it looks for a coarser representation T that retains less of X 's detail. How much detail T retains about X is measured by the mutual information $I(X;T)$ ⁶⁶. A large $I(X;T)$ means T still tells you a great deal about X , while a small one means T has thrown most of X 's specific information away. Compression is thus the pressure to make $I(X;T)$ small. On the other side it asks for relevance preservation. T should remain informative about some downstream variable Y the variable one actually cares about. The information T retains about Y is measured by another mutual information, $I(T;Y)$, and the pressure here is to make this one large. The two pressures pull against each other because the easiest way to compress X is to discard its detail indiscriminately, which inevitably also discards the part of X that mattered for Y . And the easiest way to preserve information about Y is to keep X in its entirety, which is naturally no compression (at all). The Lagrangian writes both pressures into a single function whose

⁶⁶ As per section 1.1.2

minimisation finds the best trade-off between them, the most compressed T that still preserves a specified amount of information about Y . A control parameter, conventionally written β , sets the rate at which the two pressures are weighted against each other, with a small β leaning toward compression and a large β leaning toward relevance preservation (the optimal T shifts as β varies along this trade-off).

For most values of the control parameter, the optimal representation shifts smoothly, a slightly larger β yields a slightly more relevance-preserving and slightly less compressed T . At certain critical values of β , however, the optimal representation undergoes a bifurcation, a qualitative change in which T acquires a new cluster, or an existing cluster splits, or some other discrete change in the structure of the optimal representation occurs (Tishby et al., 1999). Bifurcations can be understood as the IB analogue of phase transitions, and Tishby (et al., 1999) help in identifying the natural levels of granularity in the joint distribution of X and Y . Like phase transitions, their location is not chosen by the modeller but it is a property of the joint distribution itself.

Where phase transitions require a physical free energy and a parameterised constraint structure, the information bottleneck requires only a joint distribution and a relevance criteria. The bifurcation points of the information bottleneck curve identify configurations that compress the fine-grained variable into stable macrostates with respect to the relevance variable, that is, configurations whose macrostate structure is robust to small changes in the compression-relevance trade-off. These are the configurations that, in this framework, count as having achieved persistence at their level of description.

Two complementary formalisations, one criteria

The two formalisations above are complementary. Phase transitions ground the criteria level viability at the physical and chemical level, given free energy direct thermodynamic meaning. Information bottleneck bifurcations ground it at any level where a joint distribution and a relevance variable can be specified, which is to say at any level the framework cares to address. What both share, and what matters most for this thesis, is the absence of a free threshold parameter. The selection threshold is thus a property of where the constraint family makes the free energy non-analytic (in the phase-transition case) or where the information-bottleneck Lagrangian undergoes a bifurcation (in the IB case). A configuration either does or does not sit at a stable region of its constraint landscape, which can in principle be checked against the actual structure of the joint distribution.

1.3.3 Singularity-nodes

Both formalisations describe the same kind of phenomena, a configuration at which the constraint family undergoes a qualitative change. In the phase-transition formulation that change is a non-analyticity of the free energy with respect to a control parameter; in the information-bottleneck formulation it is a bifurcation of the optimal representation with respect to the compression-relevance trade-off. The thesis refers to call any such configuration a *singularity-node*⁶⁷.

⁶⁷ "Singularity" carries baggage from general relativity, the technological-singularity / AI tradition (as it relates to AGI), and from complex-systems writing on tipping points. While what I mean relates to none of those, it was directly inspired by their phenomena type. A singularity-node, throughout this thesis, is the statistical-mechanical sense of singularity — a non-analyticity of the free energy of a constraint family, or equivalently a bifurcation of an optimal representation under an

A singularity-node is the precise *location in configuration space* where one of two events occurs: one, the free energy of the constrained ensemble fails to be analytic in its control parameter; or two, the optimal representation in the information-bottleneck Lagrangian undergoes a bifurcation. Either event is intrinsic to the joint distribution that describes the system's components. Whether a given configuration is or is not a singularity-node is a fact about the structure of the joint distribution, checkable in principle and, in many cases, checkable in practice.

The role of singularity-nodes in this framework is henceforward threefold. First, identify where in configuration space level-transitions occur. The intuitive talk in earlier systems-theoretic literature about systems "forming" at certain points along their constraint trajectories becomes precise once the forming-point is identified with a singularity-node⁶⁸. With the post-singularity macrostate structure being what registers at the higher level as "a system having formed". Second, they ground the integrity criterion of section 1.3.1. A configuration sitting at a stable phase on the post-singularity side of the constraint landscape is robust by construction, since small perturbations leave its macrostate structure intact, and this robustness is what integrity, in the sense used here, formally measures. Third, they make the bottleneck concept of section 1.2.1 properly continuous with the selection mechanic of this section. Bottlenecks are the configurations through which many independently-converging paths happen to pass, with singularity-nodes being the configurations at which the constraint family consolidates into a stable post-transition phase. The two overlap often, because the configurations that are reused are typically the ones that sit at stable post-singularity phases of the constraint family they participate in.

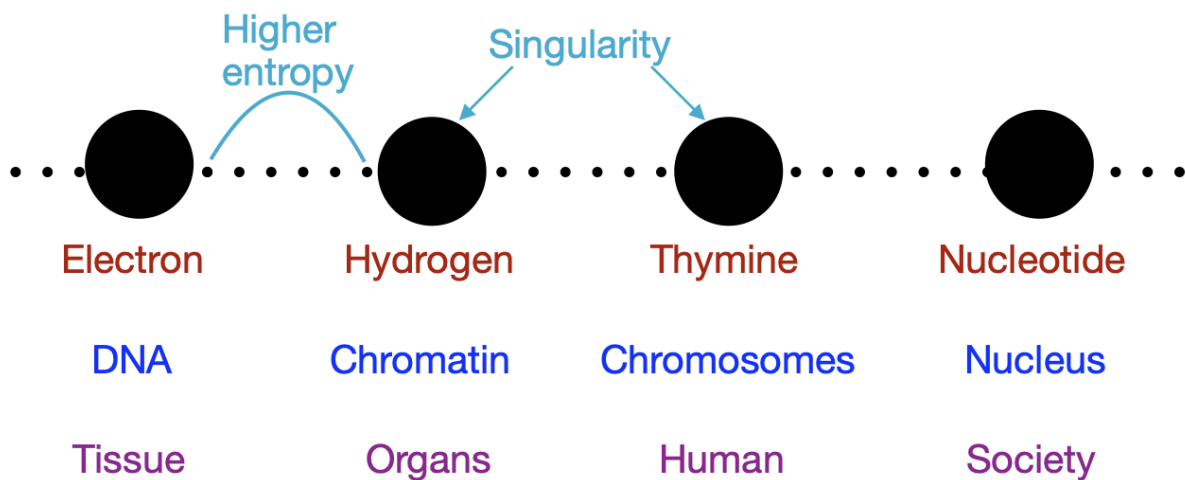


Figure 3: Singularity-nodes across scales of constraint families. Each row traces a sequence of stable post-transition configurations along an assembly path: atomic and molecular (red), genetic (blue), biological and social (purple). Each black circle marks a configuration sitting at a stable phase of its constraint family; each "Singularity" label marks a singularity-node at which the prior alphabet of microstates has just collapsed onto the post-transition macrostate. The "Higher entropy" label indicates the higher-entropy region of a constraint family between two singularity-nodes, before the next non-analyticity

information-bottleneck Lagrangian — applied to the joint distribution that describes a system's components. The "node" half of the term (although I will often throughout this thesis describe the object simply as a *Singularity*) marks that what is being picked out is a discrete location, both in the configuration space the constraint family lives over and in the hierarchy of levels the framework constructs from those constraint families.

⁶⁸ Or in the sense of section 1.1.3 a system

prunes the joint distribution onto a smaller, more concentrated alphabet. This figure is illustrative and naturally non-exhaustive, with each row corresponding to many more singularity-nodes than shown here, and the parallel layout is meant to convey that the same selection mechanic of path-reweighting toward stable post-transition phases operates at every scale of description.

1.3.4 Selection as path-reweighting

Combining the historical picture of section 1.2 with the fitness criteria above yields the selection mechanic I've begun this section with. An assembly path is a sequence of constraint applications that produces a stable configuration from an initial alphabet. The space of such paths is the assembly contingent space of Sharma et al. (2023). Across this space, paths are reweighted by the persistence-and-propagation properties of their configurations. Paths that stabilise in robust configurations contribute more probability mass to the realised assembly distribution than paths that terminate in fragile ones, because the robust terminations are the ones that participate in subsequent assemblies.

The realised distribution over assembly paths departs from the prior distribution, which would be uniform over the space of physically possible paths, in proportion to the integrity and copy number of each path's terminal configuration. This departure is negentropy⁶⁹, that is, a single distribution's departure from uniformity, applied here to the space of assembly paths⁷⁰. Selection is the process that produces this negentropy on the assembly-path space.

Two consequences follow for this thesis. First, the structure of the realised assembly graph is not the structure of the combinatorially possible one. It is heavily biased toward paths through robust intermediates. The bottlenecks identified in section 1.2.1, the singularity-nodes introduced in section 1.3.2, and the selection mechanic just stated are three descriptions of the same underlying fact, viewed from three different vantage points. Bottlenecks are the structural signature of selection in the realised assembly graph (graph-theoretic vantage). Singularity-nodes are the configurations at which the constraint family has consolidated into a stable post-transition phase (constraint-distribution vantage). Selection is the process that reweights probability mass toward those configurations and the paths that reach them (process vantage). Reading the framework with all three available is what lets later chapters connect a structural feature of an assembly graph to a phase-transitional fact about the constraint family that produced it. Second, what counts as robust is itself level-dependent. A configuration that is robust at one level of description may be fragile at another, and the level at which the fitness criteria is applied determines which paths get reweighted upward. A protein conformation that is stable in cellular conditions may be unstable in extracellular ones, while an institutional rule that is robust in one political configuration may dissolve in another, and so on. The same selection mechanic operates at each level using the level's own integrity and propagation criteria, with each level's singularity-nodes located in that level's own constraint family.

1.3.5 Next stop, the *reductionist substrate*

⁶⁹ As per section 1.1.2

⁷⁰ As opposed to the space of microstates

"Unity of Science is granted, not by a utopian reduction of all sciences to physics and chemistry, but by the structural uniformities of the different levels of reality."

(Bertalanffy, 1969, p. 88)

Finally circling back to close the question opened up at the preamble, to which *reductionist substrate* to commit to. I'll remind the reader that my commitment was to identify a *substrate* that was substrate-independent, monist, and capable of handling both the ontological critique (Nagel, 1998) and the downward causation critique (Campbell, 1974) without either denying the higher-level phenomena those critiques rest on or conceding the lower-level substrate the reductionist commitment requires.

Consider the example of domestication. The Ontological critique, as Nagel (1998, p. 5) put it, holds that some physical phenomena *"even though they can be explained in terms of principles that fit their specific features, simply do not have an explanation at the ultimate level"*. No description in terms of particle physics would aptly captures what domestication is⁷¹, because the variables that are the diagnostic characteristics of the phenomena depend on states that aren't available at the particle level, and can only be found on higher assembly indexes.

The ontological critique concludes that there is no *ultimate-level explanation*, at least not on domestication or markets, and the downstream implication to reductionism is that a framework anchored in an *ultimate substrate* does not hold. However my framework developed above understands that ontology is *level-relative*, and that the level at which a phenomenon acquires the variables relevant to its content is not a matter of analytical preference but of where the *singularity-nodes* of the relevant constraint family actually lie. Domestication is a stable post-singularity phase of the constraint family that governs interactions among human populations, animal populations, food-production strategies, and selection pressures across generations. That constraint family is non-smooth at certain configurations, such as agricultural intensification, crossing singularity-nodes that produce stable post-transition phases recognisable as domesticated species, with characteristic morphologies, behaviours, population dynamics, and feedback loops with their human hosts that no pre-singularity description anticipates (Driscoll et al., 2009). Domestication is the post-singularity macrostate, the stable phase that the constraint family at that level admits.

The same can be handled across further classical examples. Such as *temperature*, a stable phase of a molecular constraint family that becomes meaningful once thermalisation has occurred, but is not a property any individual molecule possesses, given that there is no temperature of one molecule, only a temperature of an ensemble whose constraint family has crossed into a thermalised phase. Or *markets*, stable phases of an economic constraint family that become meaningful once enough agents, transactions, and informational channels are coupled to cross the non-analyticities that produce price equilibria, but are not properties any individual transaction possesses, given that there is no market of one trade, only a market of an ensemble whose constraint family admits price-equilibrating dynamics⁷². Each is real at its level. Thus dissolving the critique in the form originally posed. The reductionist substrate — information as departure from maximum entropy on probability spaces — is preserved at every level the framework touches, including the levels at which "domestication" or "markets" are the natural variables. What is rejected is the assumption

⁷¹ "Explanations at higher levels often remain practically preferable and for many purposes indispensable" (Nagel, 1998, p. 4).

⁷² This will be further developed later on the thesis

that *the* or a “ultimate level” is a fixed level of explanation. As Morris (2015, p. 9) puts it, “*the question we should be asking is not whether a historian, an anthropologist, or a sociologist is being reductionist (...) but what level of reduction is required to resolve the problem being posed*”. The framework above gives a clear answer: the appropriate level is the one whose constraint family is stable between the relevant singularity-nodes for the phenomenon at hand.

The Downward Causation critique is the converse problem. The apparent incongruity of having structures “above” the reductionist substrate causally affecting their constituents. Campbell's (1974) provides an example with soldier ants, whose jaw morphology is so specialised for the colony-level function of piercing enemy termites that the soldier cannot feed itself and must be fed by other workers. How can morphology, molecular and atomic structures, be influenced by a colony-level role, how can the colony level act on the molecular level without violating physical laws? This framework's answer is that the colony level does not *act on* the molecular level in any sense that requires new causal forces. What it does is reweight the probability mass over molecular configurations. The constraint family governing ant development is parameterised, in part, by colony-level variables — caste-determining hormonal levels, social-feedback signals, kin-recognition cues — whose values are themselves set by the colony-level state. Different colony-level states produce different parameter values, and different parameter values place the developmental constraint family at different points in its phase diagram, with different post-singularity morphological phases as the stable outcomes. The molecular biology of the soldier ant is fully determined by molecular-biology laws, what the colony supplies is the parameter values and boundary conditions under which those laws operate. A different colony-level state would have placed the same molecular-biology laws in a different region of their phase diagram and produced a different morphology, and this is what is observed.

Formally, this is the path-reweighting mechanic of section 1.3.3. Higher-level constraint families select among lower-level paths by determining which terminal configurations sit at stable post-singularity phases. Mathematically, the move is from an unconditional distribution $P(X)$ to a conditional distribution $P(X | Y)$, where Y is the higher-level state, and conditional distributions are not new physics, they are the standard apparatus by which a joint distribution decomposes (section 1.1.4). Downward causation, in its non-mysterious form, is constraint propagating downward as conditioning, and singularity-nodes at the higher level register at the lower level as boundary-condition shifts that reweight the lower-level phase diagram. The same structure handles cases the broader thesis develops further on, such as sieve tube cells in plants⁷³ losing their nuclei, vacuoles, and most of their organelles during differentiation because the plant-level transport function reweights which cellular configurations are stable in mature phloem tissue, the molecular-biological laws governing organelle maintenance still hold, but the parameter values they operate under are set by plant-level signals that select for hollow conducting cells. The mechanism is the same throughout, higher-level constraint families project downward as conditioning on lower-level distributions, and the realised lower-level distribution is the result.

While the Ontological critique reads it upward — in that at certain configurations lower-level constraint families consolidate into stable phases whose properties cannot be read off the lower-level alphabet without going through the singularity-node — the Downward Causation critique reads it downward, i.e., those same higher-level phases, once

⁷³ Further developed in section 2.2

consolidated, set parameter values and boundary conditions that reweight the lower-level distribution, and this reweighting is conditioning. The upward flow of assembly into stable post-singularity phases is matched, on the same singularity-nodes, by a downward flow of constraints acting as informational filters that reweight the microstate distribution at the level beneath. With this mechanic in place, both critiques resolve into the same fact viewed twice with the path-reweighting mechanic that ties the two influences together.

Therefore, the *reductionist substrate* this thesis commits to information. The measurable departure of a joint distribution from its maximum-entropy reference, on whatever probability space the level of inquiry requires. Its highest expressions are viability (which configurations get reused), assembly bottlenecks (where in the assembly graph the reuse concentrates), and singularity-nodes (where in the parameter space the consolidation that makes a level happens). Despite formalised through different *tools*, Bertalanffy (1969, p. 75) inspirational intuition remained whole, with [levels'] "*hierarchical structure and combination into systems of ever higher order, [being] characteristic of reality as a whole and of fundamental importance especially in biology, psychology, and sociology.*"

Chapter 2

Non-equilibrium Integrity, Biology and Cognition

The epistemology established from first principles in the first chapter, proves the possibility of anchoring information as a reductionist substrate, from which to examine functionally processes in light of their material-energetic-informational roles in gaining information: either from pruning lower level alphabet states (non-compatible with higher level assemblies), or from pruning generative models non-compatible with homeostasis⁷⁴. The concept of *processes* (and by extension *functionalism*, as developed in 2.7 below) here is key. It is my belief that the failure to establish the primordial principles and mechanics of *processes* at different scales is the root cause for silos, disparate jargon and methodologies across the Biological, Cognitive and Political disciplines.

2.1 Acquired tastes: from constraints to operations (Cognition as persistence in non-equilibrium)

In the history and memory (as per Assembly Theory terminology) of a given system's possibility space, there are assembly routes that can lead to a phase transition in which the system's lineage goes from existing in a state of equilibrium to non-equilibrium. This is an assembly bottleneck, and only systems capable of actively maintaining their integrity will continue exploring further potential assembly pathways. This is where the principles of *Basal Cognition* become integrated⁷⁵, given that the whole architecture of cognitive solutions emerge from the moment such systems' historicity pass the assembly bottleneck of finding active solutions for integrity.

The critical aspect of this phase transition regards a distinction in the nature of constraints. While Chapter 1 established how passive constraints reduce uncertainty and enable higher-order assemblies, there exists a fundamental threshold beyond which passive constraints become insufficient. Passive constraints — such as chemical bonds, geometric compatibilities, or thermodynamic favorability — can maintain relatively simple assemblies in or near equilibrium states. However, as assembly index increases and systems venture further from equilibrium, these passive mechanisms prove inadequate against the pull of entropy. Consider the fundamental difference between a crystal formation and a living cell. A crystal grows through passive constraint satisfaction: molecular components arrange themselves according to fixed geometric and energetic rules, achieving a stable, low-energy configuration. The crystal's structure is maintained by static forces, such as electromagnetic interactions between its constituent particles (mutual sub-system restraints) that require no ongoing energy expenditure. In contrast, a cellular system exists in a state of dynamic equilibrium, continuously exchanging matter and energy with its environment while maintaining actively its integrity (Schrödinger, 1944; Nicolis & Prigogine, 1977). The cell cannot rely on passive constraints alone because its non-equilibrium state would rapidly decay without active intervention.

⁷⁴ In both cases, information is gained as the measurable departure of a joint distribution from its maximum-entropy reference, instantiated through the three complementary registers developed in 1.1.2 propagated along the path-dependent assembly routes of 1.2, and consolidated at the singularity-nodes of 1.3

⁷⁵ As maintaining integrity in non-equilibrium systems constitutes homeostasis, as it will be explored soon in this section

This transition from passive to active constraints represents a qualitative shift in the nature of assembly constraints. Active constraints are operational. They involve dynamic processes that continuously monitor system state, compare it against preferred configurations, and execute corrective actions when deviations are detected. In a system's historicity, once it enters into a non-equilibrium state, the system either drifts back into the assembly memory state(s) in equilibrium, vanishes from the assembly lineage altogether, or it self-maintains in a non-equilibrium state. This enforces a bottleneck in assembly history ensuring that non-equilibrium systems' assembly ought to exhibit minimal homeostatic solutions to venture into further non-equilibria assembly spaces. Only those constraint configurations capable of transitioning from passive to active operations can persist beyond this critical threshold. As per 1.3.4, the realised distribution over assembly paths past this bottleneck concentrates exclusively on those whose terminal configurations possess the integrity-and-propagation properties required for non-equilibrium persistence. Thus Cognition is what lets a non-equilibrium system persist, and instead of being understood as achievement reached at some threshold of neuronal networks architectures, it ought to be seen as admission into post equilibrium assembly indexes that actively pursue integrity.

In assembly history, the emergence of active constraints can thus be understood as a specific type of memory, operational memory, that encodes not just successful assembly pathways but dynamic procedures for maintaining assembly integrity. Unlike passive memory, which simply preserves structural templates, operational memory maintains procedural knowledge about how to respond to environmental perturbations. This represents a qualitative leap in informational capacity, as the system transitions from merely storing "what worked" to encoding "how to make it continue working."

Thus, informational features such as distinguishing *environment* versus *system* or "*ensuring certain parameters stay in specific ranges*" (Levin, 2019, p. 13) are informational cognitive characteristics upon which *life* itself has dependencies. That is, the most primitive feature of a free energy minimising system are *preferences*⁷⁶, i.e. the fact that some states of the world are better for facilitating its integrity and diminishing *surprise* than others. From such *preferences*, learning from positive and negative reinforcement opens up, leading to an explosion of computational and behavioral capabilities that further expand the assembly pathway of non-equilibrium systems towards higher complexity. A preference here thus has nothing to do with conscious desire, being just a bias in the system's distribution over possible states. Preferences can be understood as non-uniform prior distributions $P(s)$ where certain states s have higher probability than others, due to these priors emerging from the system's assembly history of configurations that have proven viable for maintaining integrity. Each preference embodies accumulated constraints from successful survival episodes, functioning as a prior belief about which states are likely to maintain system integrity. For instance, a bacterium's preference for nutrient-rich environments emerges from evolutionary assembly history where ancestors who exhibited such biases were more likely to maintain their organisational integrity and continue their assembly lineage. Preferences thus serve as the bridge between historical constraint accumulation and active constraint implementation. They provide the evaluative framework necessary for active inference, the mechanism by which systems can assess whether their current state aligns with integrity-maintaining configurations and take corrective action when misalignment is

⁷⁶ "homeostatic persistence is the origin of cognitive goals" (Levin, 2019, p. 13)

detected. This approach provides a framework to how goal-oriented systems maintain their structural and functional integrity in the face of environmental interactions⁷⁷.

Thus — past that assembly informational bottleneck — each system, no matter how rudimentary, is actively driving toward some target conditions that it *cares about* (its valence, or preferences — certain configurations, arrangements, locations are better than others to the degree that it allows maintenance, or decreases free energy, better than others) to maintain integrity⁷⁸. The system *cares* in the sense that it exhibits systematic behavioral patterns that favor certain outcomes over others, not through conscious intention but through ingrained bias structures that emerged from assembly selection pressures. This caring manifests as operational constraint satisfaction, i.e., the system continuously working to maintain states that align with its encoded preferences. This framework was inspired by *Basal Cognition* understanding of *agents*⁷⁹ as endowed with minimal homeostatic loops stemming from volition and preference — even if that preference is as simple as “nutrient rich environments over depleted ones” or “establishing relations with kin over non-kin”⁸⁰ — and thus by extension granting them with *agent level goals*.

Whereas the previous section showcased *constraints* in general physical systems, Active Inference applies restrictively to higher level systems that must actively maintain their form and function despite continuous interactions with a changing environment⁸¹. So, the question that it addresses is, how do cognitive (functionally adaptive towards a goal state, in the broadest sense homeostasis) systems *persist* over time (that is, maintain their structural integrity while continuously exchanging matter and energy with the environment)⁸². Active Inference represents the operationalisation of preferences into dynamic constraint management. Active inference systems proactively shape their alphabet of states to align with their preferred states. This represents a fundamental shift from reactive to proactive constraint satisfaction. The key mechanism underlying Active Inference is the concept of a generative model, an internal representation of how the world works that enables the system to predict the consequences of different actions (Parr et al., 2022). This generative model is not a passive mirror of reality but an active, preference-laden construction that biases perception and action toward integrity-maintaining outcomes. *Escherichia coli*, for example, move toward higher concentrations of nutrients, and once the concentration decreases, they undergo tumbles to reorient themselves until they find a favorable direction, a cognitive search strategy in navigating the nutrition acquisition problem-space. “*The overall direction of bacterial swimming is subject to sensory-based cellular control by altering the direction of the flagellar motors in response to a signal transduction system linked to spatially organized and tunable receptors at the cell surface. Net positive chemotaxis towards a localized nutrient source, for example, occurs when the receptors send signals for CCW rotation (runs) as long as the nutrient concentration is increasing and signals for CW rotation (...) when the nutrient concentration falls*” (Shapiro, 2021, p. 135). This example illustrates the

⁷⁷ “All of these responses act to restore an overall homeostatic setpoint, i.e. an expected nonequilibrium state; hence they can all be viewed as acting to minimize environmental variational free energy or Bayesian expectation violation” (Fields, Glazebrook, & Levin, 2021, p. 10)

⁷⁸ As Michael Levin puts it: “I propose that the ‘atom’ of this cognitive hierarchy is homeostasis” (Levin, 2019, p. 13)

⁷⁹ For a definition and understanding of what this thesis conceptualises as *agents*, see section 2.2

⁸⁰ “One of the most important bacterial cognitive capabilities in multicellular communities is distinguishing clonally related cells (self) from unrelated cells, either of the same or different species” (Shapiro, 2021, p. 136)

⁸¹ “The embodied agent has a ‘do or die’ to-do list to maintain its system boundaries, or more comprehensively, to survive and thrive (...). This set of agentive processes can be understood as an active engagement in a homeostasis/allostasis dynamic, which more broadly still, can be regarded as adaptive behavior.” (Linson et al., 2018, p. 19)

⁸² “[we should reject the representational view of cognition altogether. Instead,] cognition should be viewed as a process of self-organization among the components of the (...) system performing the cognitive activity. (...) On that view, cognition is a homeostatic and allostatic process of attunement to cope with environmental perturbations” (Constant et al., 2021, p. 2)

architecture for active constraint management. The bacterium's "generative model" consists of simple rules: "moving toward increasing nutrient concentration promotes integrity; moving toward decreasing concentration threatens integrity." These rules, encoded in molecular machinery, enable the bacterium to actively shape its trajectory through nutrient space instead of passively drifting. The tumbling behavior represents active exploration when current actions fail to align sensory input with preferred outcomes. The bacterium uses current sensory data to update its beliefs about the likelihood of different spatial locations containing nutrients, then acts to test these predictions. This inference loop creates a dynamic coupling between internal model and external environment that maintains the system far from equilibrium through continuous active adjustment.

Through Active Inference one can extend these principles to any dynamic adaptation and persistence within systems at a non-equilibrium state level. Active Inference shows how systems on *alphabet states* which are non-equilibrium actively seek to reduce uncertainty iteratively, adjusting both internal models and external actions (Parr et al., 2022). Active Inference as an intra-level dynamic framework deepens and complements the previous inter-level one (of 1.1.4 and 1.3): not only do higher-level structures constrain lower-level states, but agents embedded in these structures continually refine their 'generative models' to maintain their viability and integrity within those states (Parr et al., 2022). Active inference posits that cognitive systems achieve such *persistence* by minimising surprise (or prediction error) through reciprocal action-perception loops. Each action shapes the system's future sensory states, while each observation updates internal beliefs⁸³ — or generative model. It does so by conceptualizing cognition as an engine that integrates into prior *beliefs* new *evidence*, to refine its internal generative model, in a Bayesian manner⁸⁴. Thus environmental input — or *perception* — functions as hypothesis testing, refining probabilistic representations to minimize prediction errors as to the output's — or *action* — effect⁸⁵. When the system encounters data that conflicts with its preferred states, this generates "surprise" or prediction error. The system then faces two options, either update its internal model to better accommodate the unexpected data, or take action to change the environment (or its relation to it) to better align with its preferred states. The balance between these two strategies depends on the system's confidence in its current model and the cost of environmental modification. Persisting as a distinct entity in a changing world means resisting the natural drift toward higher entropy, which — as attested in the previous sections — can be framed as reducing uncertainty. A system capable of modelling the environment with probabilistic assumptions (by Bayesian inference), about the hidden states generating input data, can refine it so as to iteratively reduce errors in its model. Active Inference formalises the reduction of uncertainty — that is, the surprise between the expected outcome according to the current world model and the actual environmental response as recorded via input — as the minimisation of *variational free energy*. Free energy is the same object at 1.3.2 used to locate singularity-nodes in physical constraint families, and here being minimised dynamically instead of mapped statically, exposing the deep continuity between the static phase-diagram view of Chapter 1 and the dynamic active-inference one.

⁸³ "Essentially, for each action-perception cycle, the environment sends an observation to the organism. The organism uses (an approximation to) Bayesian inference to infer its most likely hidden states. It then generates an action and sends it to the environment in an attempt to make the environment less surprising. The environment executes the action, generates a new observation, and sends it to the organism. Then, a new cycle starts." (Parr et al., 2022, p.24)

⁸⁴ "Active Inference subsumes and extends these ideas by treating not just perception but also action, planning, and learning as problems of (Bayesian) inference" (Parr et al., 2022, p. 15)

⁸⁵ "Perception is not a passive outside-in process—in which information is extracted from impressions on our sensory epithelia from "out there." It is a constructive inside-out process—in which sensations are used to confirm or disconfirm hypotheses about how they were generated" (Parr et al., 2022, p. 16)

As Clark and Friston (Linson et al., 2018, p. 14) put it — “*FEM* [Free Energy Minimization] is a reduction of uncertainty, i.e., an increased probability of picking out the relevant signal from the noise” — right before tying up the concept in accordance to the first principle from which the *measurable departure-from-maximum-entropy* developed in 1.1 was derived: “By analogy to physics, this quantified uncertainty is termed Shannon entropy. Higher Shannon entropy reflects a greater uncertainty in picking out the relevant information, so informational FEM amounts to improving the identification of the relevant information.”

Active Inference goes beyond the recognition that perception and action have the same (inferential) nature. It also assumes that both perception and action cooperate to realise a single objective, or optimise just one function instead of having two distinct objectives, as more commonly assumed. In the Active Inference literature, this common objective has been described in various (informal and formal) ways, including the minimization of surprise, entropy, uncertainty, prediction error, or (variational) free energy. (Parr et al., 2022, p. 24 - 25). This unified objective being the operational implementation of maintain configurations that support continued existence, via operational constraints to states compatible with the systems priors (*preferences*). All perceptual and active processes serve this meta-constraint, with preferences providing the evaluative criteria for determining which configurations merit preservation and which require modification. As per 1.1.2, the agent is acting so as to keep the negentropy of its own state distribution high — concentrating probability mass on the small alphabet of microstates compatible with its integrity — and the mutual information between its sensory states and its hidden states tightly coupled to its preferences. The constraints the agent imposes on itself thereby propagate the same kind of state-space compression that 1.1 located in physical constraint families, now realised through an internal generative model and not through chemical bonding rules or quantum selection rules. Active Inference show how systems — by reducing surprise — minimise entropy and gain information via pruning the space of perception-action inferences.

Thus — whereas information at the *level-to-level* order quantifies how much *constraints* have reduced the state space towards lower-entropy higher-level emergence — at the cognitive system order it quantifies the degree to which the system’s generative model explains its sensory input. Thus, for systems that must actively maintain their form, minimising free energy ensures internal consistency and alignment with environmental states, ensuring that the system navigates its high-dimensional state spaces so as to remain in preferred (low-surprise) regions. In a sense, relating to the approach above, the constraints the system imposes on itself — the *optimistic biases* encoded in their generative models as fault preferable states — perform a role akin to the *messages*, specifying which states are viable in their configurational homeostatic space⁸⁶. The integration of passive constraints (Chapter 1) with active operations (this section) reveals a hierarchical architecture of system maintenance. At the foundational level, passive constraints establish the basic viability conditions, the physical and chemical rules that determine which configurations can exist. At the operational level, active constraints implement dynamic maintenance procedures that keep the system within viable regions of its state space despite environmental perturbations. This hierarchical structure explains why active

⁸⁶ “preferred states (or observations) can be specified as the priors of the model—which implies that the model implicitly assumes that its preferred (prior) sensations are more likely to occur (i.e., are less surprising) if it satisfies the criteria for existence. This means it has an implicit optimism bias. This optimism bias is necessary for the agent to go beyond the mere duplication of external dynamics to prescribe active states that underwrite its preferred or characteristic states.” (Parr, Pezzulo, & Friston, 2022)

inference systems can achieve such remarkable stability in non-equilibrium conditions. The active layer continuously monitors and corrects deviations from preferred states, while the passive layer provides the stable foundation upon which these corrections operate. Neither level alone would be sufficient, as passive constraints without active maintenance lead to equilibrium drift, while active operations without passive foundations lead to unsustainable energy expenditure.

By framing this free energy minimisation heuristic as hypothesis testing, cognitive systems act not just to receive informative data, but to generate it via prediction error (the discrepancy between expected and observed sensory data). By highlighting where the generative model fails to capture environmental patterns, it enabled resolving these errors through updates in beliefs or through environmental interventions. Given that each resolved prediction error corresponds to a reduction in uncertainty, one could say the system has gained a “bit” of information, reducing future prediction errors. From this perspective, cognitive agents are not passive recipients of signals but active shapers of their own data streams, continually rearranging the *alphabet* of possible sensory states to align with their generative model’s constraints. This resonates with the showcased at 1.1⁸⁷ now extended into the temporal domain of action and perception cycles. This active shaping of data streams represents a qualitatively new form of constraint propagation, no longer simply inheriting constraints from assembly history, the system becomes an active participant in *constraint generation* and *refinement*. Through the prediction error minimisation process, successful constraint patterns are reinforced while unsuccessful ones are modified or abandoned. This creates a feedback loop between constraint effectiveness and constraint evolution that drives increasing sophistication in system maintenance capabilities. Active Inference thus operationalises cognition while remaining anchored to the same reductionist substrate adopted in 1.3.4, of information as the measurable departure of a joint distribution from maximum entropy.

Thus, for this thesis — either when analysing within the scope of a single cognitive system (being this a low or high order one), or considering level to level emergence — *information* (as departure from maximum entropy in the joint distribution over a system’s accessible microstates) emerges as the scale-free reductionist substrate. This will serve as the conceptual backbone for examining cognitive and political concepts alike.

Adopting this assembly and informational perspective has several downstream implications I would like to mention (although the argument for this subsection has already been delivered) since it adds to the non-biocentric stance of this thesis, and clarifies the epistemological landscape treated towards understanding Governance Assemblies as cognitive. While steering away from what is typical in biology’s lexicon, homeostasis — maintaining the internal milieu of a system by leveraging free energy to remain far from equilibrium while adjusting and adapting structure and action as the environment fluctuates via feedback loops — maintains its role, although shifting from metabolic to informational in a definition meant to accommodate this framework. Beyond focusing only on chemical or energetic balance, one can thus understand homeostasis as the maintenance of preferred information patterns, configurations of states in assemblies that align with integrity-supporting priors. This reframing extends homeostatic principles beyond biological systems to any assembly capable of active constraint management.

⁸⁷ Information as departure of a distribution from its maximum-entropy reference

This shift is essential to appreciate how its role overlaps with the system's ability to *read*, anticipate, learn-from and adapt to its surroundings, stemming directly from its striving towards integrity. Each of these operations correspond to the ability to maintain a circuit of input and output while resolving a problem space. In maintaining and updating this circuit, a system is constructing and refining its internal generative model(s). As a result, cognitive attributes become rooted in the mere aspect of maintenance. Thus, homeostasis is no longer just a chemical or physiological activity — but a recursively updated, historically informed cognitive process — an embodied form of problem-solving irrespective of its medium. But also, recognising homeostasis as intrinsically cognitive elevates the concept beyond mere metabolic stasis. It infuses it with the dynamism of inference, strategy, and adaptation. When seen through the lens of this framework, homeostasis becomes a universal signature of systems that can integrate sensory signals, model their environments, harness free energy, and preserve themselves against the drift toward thermodynamic equilibrium. By extension, what was once considered a strictly biological phenomenon morphs into a principle with implications for any non-equilibrium self-organizing system. It effectively closes the gap between life and cognition at the fundamental operational level, suggesting that wherever one finds systems actively stabilising themselves far from equilibrium, one also finds the rudiments of cognitive function. As Levin (2019) puts it in his paper *"The Computational Boundary of a Self"*, *"cognition, like "organism status" (...) is not a binary capacity that exists in higher organisms only. (...) it is fundamentally incorrect to view functional capacities such as memory, prediction, goal-directedness, etc. as entirely new features appropriate only to advanced forms of life (...). failure to appreciate cognitive processes in a system that would have otherwise improved prediction and control is a "neganthropomorphic fallacy" – a type 1 error (...). Thus, here it is assumed that there is no such thing as magical "true cognition" that is the province of humans"*.

As per the last point, whenever a non-equilibrium system forestalls thermodynamic decay, preserving its structural integrity through the active reduction of free energy gradients — a process inherently reliant on adaptive environmental inference — it does so by acquiring, retaining and processing internal, external and generative (in terms of world model and goal-state) information. As shown in section 1.3.4, the same reductionist substrate operates across every level, while bearing the flexibility to be materialised as different types of possibility-space constraints. The consequence of this is understanding Cognitive processes (as information processing towards systemic integrity) as scale-free principles. An expression of universal processes that can operate at any level in which complexity can afford running feedback loops to track and anticipate changes so as to actively maintain structural integrity. Single cell, organism or swarm, composed by carbon-based biochemistry or otherwise. What matters is the ability to translate free energy into organised states, guided by historical information encoded within assembly pathways and adaptive structures. This new frame of reference challenges traditional conceptions of what counts as cognitive. By acknowledging its scale-free nature, biological, cognitive and psychological sciences unlock a transdisciplinary convergence with physics, information theory, and — how I argue in this thesis — law and political science, that enable an holistic study to the fundamental processes of systems that persist and flourish within the dynamic theaters of a non-equilibrium existence.

2.2 Motivated collaborators: integrity all the way down

The transition from passive to active constraints established in the previous section reveals a fundamental principle: non-equilibrium systems must become cognitive to persist. As seen above, Cognition intrinsically entails the capacity to evaluate options and make choices⁸⁸. This is best evidenced when examining cellular systems. It helps to follow the framework above and regard a cell as an historically contingent, information-driven⁸⁹ set of constraint building (in relation to constituent statespaces) operating within the assembly problem-space of maintaining integrity at a nonequilibrium assembly index coordinate. This characterisation positions cells not as mechanical components but as cognitive agents navigating complex problem spaces. Or, as Levin (2023, p. 1) puts it, “*cells (...) are not passive components (...) they have numerous capabilities for behavior*”.

Cells operate along two fundamental axes of constraint that shape their cognitive architecture. First, they constrain the available state-space inherited from lower-level components, the assembly index of any cell emerges from the mutual restrictions imposed by its molecular constituents, producing the positive total correlation that section 1.1.3 used to define a system, creating an entropy-reduced configuration space where only certain microstates remain compatible with cellular integrity. Second, cells continuously refine their predictive models of environmental interaction, updating or discarding generative models that fail to maintain homeostatic parameters. Models that inadequately predict temperature gradients or nutrient distributions, for instance, either undergo revision through active inference or result in cellular failure that eliminates both the cell and its flawed model from the assembly landscape. Whether constraining physical microstates to viable configurations or updating predictive models to better anticipate environmental dynamics, cells gain information by collapsing vast possibility spaces into narrow bands of functional states. This reduction - from the theoretical totality of possible configurations or models to the restricted subset compatible with persistence - constitutes measurable information gain.

This dual pruning process, of both physical states and cognitive models, demonstrates that cells actively participate in shaping their own constraints, not merely passively receiving them. This claim is sustained by accepting that behavioural heterogeneity under similar circumstances indicates individual evaluation and decision-making, not deterministic programming. Levin (2023) frames this plasticity as problem-solving, stating that “*cells (...) exhibit regulative plasticity (...) to adjust to perturbations such as external injury or internal modifications and still accomplish specific adaptive tasks across metabolic, transcriptional, physiological, and anatomical problem spaces*”. For example, this cognitive autonomy is demonstrated in the bistable behaviour of *E. coli*'s lactose utilisation networks, as Ozbudak et al. (2004) showed, genetically identical cells can maintain distinct phenotypic states — either induced or uninduced for lactose expression — depending on their individual history and not only depending on current environmental conditions. This hysteretic response demonstrates that cells maintain internal states that influence their decision-making. Levin (2023) extends this understanding by showing how such cellular decision-making operates according to William James's definition of intelligence as “*the ability to reach the same goal by different means*” (p. 9). This capacity

⁸⁸ As per the section above: “*When the system encounters data that conflicts with its preferred states, this generates “surprise” or prediction error. The system then faces two options, either update its internal model to better accommodate the unexpected data, or take action to change the environment (or its relation to it) to better align with its preferred states.*”

⁸⁹

for “*adaptive behavior[s]*” at “*lower level[s]*” (Levin, 2023, p. 6) represents what Levin calls “*basal cognition*”.

Each cell maintains its own generative model of environmental states, continuously updating beliefs about nutrient gradients, toxin concentrations, mechanical stresses, social signals from neighboring cells and so on. These are not simple stimulus-response mappings but sophisticated inferential processes involving uncertainty quantification, hypothesis testing, and strategic action selection. Levin (2023) provides striking examples of this cellular intelligence in planaria exposed to barium (an evolutionarily novel stressor) “*rapidly re-grow new heads that are completely barium insensitive*” by up and down-regulating specific genes to maintain physiological setpoints “*despite the inability to regulate potassium*” (p. 15). Remarkably, “*barium blockade is evolutionarily novel — planaria very likely have never encountered the exotic stressor of barium during their evolutionary history*”, yet they demonstrate “*ability to navigate physiological and transcriptional spaces to solve an entirely new problem*” (p. 15). Thus, generative models non-compatible with homeostasis are pruned, in the sense that cells' generative models that don't promote homeostasis are either updated or, leading to a non-fit cell, are extinguished along with the cell. In either way, information via cell assemblies is gained by restricting all possible microstates or all possible generative-models to a subset of much fewer viable ones.

Scaling one's proportion to the perspective of an atom, cells evidence themselves as busy consortiums in constant flux — ions traversing membranes, proteins being synthesised and degraded, metabolic pathways continuously cycling – with a recognisable identity not as single monolithic of permanency, but an informational, operational protocol enacted by ephemeral parts with function-oriented tasks that ensure the continuity of the *process*. Attesting the *Assembly* need for cognitive function immediately post the assembly *bottleneck for non-equilibria integrity*, even ostensibly “simple” cells reveal multiple layers of nested sub-processing within them, ensuring that each subunit has some internal capacity to detect and correct local errors, bridge local unpredictability and solve local issues (from damage repair, to resource pursuit) without waiting for higher-level *orders*, while actively adjusting sub-goals in real time and reducing local surprise (Michael Levin, 2019). For example — within the cell — Mitochondria act as decision-making hubs navigating metabolic problem-spaces, integrating calcium and stress signals. Through mitochondria-associated ER membranes (MAM), they control ATP production, mitophagy, or apoptosis and process information via mitophagy, mitofission, and mitofusion to maintain functional integrity, all the while responding to immune triggers, releasing mtDNA to activate the cGAS-cGAMP-STING pathway “*Once the mitochondria undergo persistent damage*”, to aid in “*microbial DNA recognition*” (Genovese et al., 2020, p.2). These nested sub-processing layers position mitochondria as autonomous agents. It is the local cunning and valence of each sub-agent cumulatively that make multi layered systems so resilient, flexible and adaptable. Levin (2023) emphasises this nested agentic architecture by stating that “*each level's components are problem-solving agents*”. Understanding this means contemplating that the evolution of multicellularity doesn't represent a loss of cellular autonomy⁹⁰ but instead an arrangement where individual cells retain their cognitive capacities while gaining access to collective computational resources⁹¹. The recursive nesting of sub-system inside sub-system is also painted by 1.1.4's grounding with each level corresponding a coarse-graining of the level beneath, and the informational quantities at each level following by pushforward from those

⁹⁰ As cellular structures still accommodate sub-cellular agency, as seen in the mitochondria example above

⁹¹ As shown in the paragraphs below

at the level below. This voluntary aspect of cellular collaboration becomes evident when examining the mechanisms through which cells coordinate their activities.

In cells, Levin (2023) argues that bioelectric communication acts as a *glue* binding together individual cellular drives towards higher order *organism* levels, being this bioelectric signaling the informational driving force scaling-up the evolution of cognition. Long before specialised neurons evolved⁹², cells' ion channels, electrical synapses (gap junctions), and voltage gradients could already coordinate a distributed computational network, which allowed for the homeostatic goal-directedness of a single cell to become informed by its neighbours' states. As Levin (2023) explains, "*All cells participate in electric networks, which allows them to compute and increase the size and complexity of simple homeostats*" (p. 12). For example, oscillatory ion waves help biofilm cells share nutrients over time (Levin 2019) – an early analog instance of multi-cellular coordination – so, in essence, single-cells that could leverage ionic currents and voltage gradients to tap into surrounding cellular states, informing their operations, had an evolutionary advantage in better handling tasks such as finding nutrients and avoiding stress over cells relying on chemistry alone. That is, the cell is individually better off by tapping into a communication network with advantages it. As Levin (2023, p. 7) illustrates, "*evolution discovered very early on, around the time of bacterial biofilms, that electric networks are also extremely convenient ways to integrate and process information across space and time, and to scale the functions of individual computing units towards a large-scale task*". Moreover, akin to the modularity offered informationally by *Assembly Bottlenecks* (section 1.2.1), Levin (2023, p. 13) documents how "*Bioelectric networks provide modularity (triggers of complex subroutines, such a simple voltage state that triggers the "build an eye" or "build a leg" subroutine)*". Furthermore, at higher-level multicellular level, the need for a medium supporting information distribution and integration over longer distances through cell networks favoured electrical solutions.

While protein structures of gap junctions and other ion-conducting pathways – that link cells into electrical networks – are directly encoded in DNA, the patterns of electrical activity function as an informational layer on top of genetics. "*DNA does not directly specify geometrical arrangements of tissues and organs*" (Levin & Martyniuk, 2018, p.76), that is, DNA alone does not contain a literal blue print of the body's shape, meaning that there exists an additional process of encoding and decoding pattern formation, a "bioelectric code" enabling the necessary coordination and communication between cells in morphogenesis — which isn't rooted in the same layer of protein instructions code — providing *instructions* and guiding articulation based on voltage, ion flux, and electric patterns (Levin & Martyniuk, 2018). In practice this means that changing a cell's electrical state can change morphological outcomes without altering its DNA. Experimental evidence (Durant et al., 2017) shows that, by inducing certain bioelectric state patterns in planarian flatworms tissue, cells could be induced to navigate a "build a head" transcriptional and morphological problem space where a "building a tail" problem space should instead naturally be, yielding a two headed flatworm system, with no DNA change, while permanently altering the worm's body pattern. This change persisted in future regenerations without further intervention, strongly suggesting an informational bioelectric memory independent of DNA⁹³. Thus Bioelectric networks serve as a medium for information, not explicitly written in the DNA,

⁹² "*It is becoming increasingly appreciated that communication via electrical processes is not unique to nervous systems, but in fact evolved continuously from far more evolutionarily-ancient properties that cells possessed long before nerves and brains evolved*" (Levin & Martyniuk, 2018, p. 78)

⁹³ The two-headed planaria kept producing two-headed offspring (when cut) not because their genes had changed – their DNA was unaltered – but because the bioelectric pattern (the target anatomical state encoded in the electrical network) had been reset and stably maintained in the tissue (Durant et al., 2017)

coordinating cells' functional articulation, and problem-space mapping across individual cognitive-light cones⁹⁴ (Levin & Martyniuk, 2018). A fun example of an individual cells navigating a problem space by leveraging the bioelectric states or its neighbours can be found in the formation of pigment patterns in zebrafish's skin, where "*the melanophore [darker pigment cell] membrane transiently depolarizes when contacted with the dendrites of a xanthophore [lighter pigment cell]. This depolarization triggers melanophore migration to avoid further contact with the xanthophores*" (Inaba et al., 2012, p. 677). This colourful morphology problem space illustrates how underneath the black and white layers there lies an individual migration, informed by communication, facilitated by bioelectricity. On aggregate, communication bridges scales – individual cells interpret voltage to locally inform their operations, and, through network scaling, tissue, organs and organisms acquire bottom-up global patterns and carry collective decisions with behaviour irreducible to a single cell's behaviour – in effect turning bioelectricity into a functional information processor sitting between genotype and phenotype as a functional *programming language*⁹⁵, and the *glue* binding cellular individual homeostatic drives into the route towards higher-level assemblies.

From the assembly, information centric lens established, the illustrations laid out above can also be interpreted so as to support the claim that bioelectric signals suppress inappropriate cellular organogenesis – ensuring that cells perform their intended *functions* instead of exploring other developmental state spaces – for example, a differentiated somatic cell in a stable voltage environment will usually stay specialised⁹⁶, but if that environment is perturbed, the cell may revert to a more primitive, stem-like program⁹⁷ or join a tumorigenic growth⁹⁸. A great contrafactual illustration of the *uncertainty mitigation* and *state space pruning* role played by bioelectric communication, comes from experiments where cells are removed from their native embryonic context – freed from the organism's normal signaling environment (thus also from its bioelectric cues) – and then are observed assembling novel, self-directed forms. Such research on *Xenobots* (*independent* frog-cell robots) (Ebrahimkhani & Levin, 2021) and *Anthrobots* (*independent* human-cell robots) (Gumuskaya et al., 2023) illustrate how cells — free from the state space pruning imposed by group cell bioelectric communication — spontaneously built new coherent structures capable of independently coming up with novel modes of behavior for locomotion and problem solving. This not only showcases the inherent flexibility and self-motivation of individual cell agents, but it also highlights the role of the whole in restricting the attainable alphabet of individual states compatible with higher level assembly, with relational restrictions being mutually imposed by same level agencies towards entropy reduction and the establishment of an *organism* singularity node. Thus, when the coordinating bioelectric

⁹⁴ The explanation, definition and utility of Cognitive Light Cones is developed at section 2.4 "A Kaleidoscope of Selves"

⁹⁵ "Expansions of the toolkit of synthetic biology will soon allow the rational top-down programming of bioelectric circuits, which will have important implications for regenerative medicine, cancer biology, and bioengineering" (Levin, 2014)

⁹⁶ "cell types have shown a strong correlation between membrane potential and differentiation, with V_{mem} becoming more hyperpolarized as cells are specified. For example, the neural crest cells of quail embryos exhibit a -35 mV resting potential early in development, but as development progresses the resting potential shifts to -55 mV" (Blackiston et al., 2009)

⁹⁷ "at the level of single cells, V_{mem} determines cell plasticity and proliferation potential. Depolarized cells tend to be rapidly proliferating and undifferentiated (e.g., embryonic, stem, or tumor cells) while terminally-differentiated somatic cells tend to be highly polarized. Importantly, cell state can be functionally altered (switched between these two classes, in either direction) by artificial change of V_{mem} ." (Levin, 2013)

⁹⁸ "Importantly, depolarized transmembrane potential is not only a marker of cancer but is functionally instructive: susceptibility to oncogene-induced tumorigenesis is significantly reduced by forced prior expression of hyperpolarizing ion channels. Importantly, the same effect can be achieved by pharmacological manipulation of endogenous chloride channels, suggesting a strategy for cancer suppression that does not require gene therapy. Together, these data extend our understanding of the recently-demonstrated role of transmembrane potential in tumor formation and metastatic cell behavior. V_{mem} is an important non-genetic biophysical aspect of the microenvironment that regulates the balance between normally patterned growth and carcinogenesis." (Lobikin et al., 2012)

network — the *cognitive glue* — is removed, cells express a wider repertoire of genome transcription, with a higher degree of flexibility driven by individual homeostatic drives. In the constraint-family understanding of section 1.3, removing the bioelectric glue is removing the parameter values that placed the cellular constraint family at its tissue-coherent phase. The cells' developmental potential then re-explores regions of the constraint landscape that were previously frozen out by collective pinning. The pruning of states compatible with organism integrity is therefore enforced at this assembly stage by a *Cognitive Glue* in the form of bioelectric communication, that propagates downward as the conditioning described in section 1.3.4⁹⁹, ensuring compliance toward the coherent function of a whole organism assembly.

The Xenobot and Anthrobot experiments — along with the illustration of how communication enhances individual cell integrity — provide the demonstration of cellular cognitive autonomy. When released from the bioelectric higher level cognitive glue that creates constraints towards organismal assemblies, cells do not simply die or revert to unicellular behaviour, but rather actively innovate, creating entirely novel morphological and behavioural solutions. This capacity for novelty, generating solutions beyond their genetic programming, establishes cells as agents as opposed to deterministic machines, having *opted* in their assembly history towards *collaboration*¹⁰⁰ because it was proven advantageous for individual integrity maintenance.

As lower level assemblies coalesce into higher order ones, downward causation in assembly can be observed as higher level systems, regardless of their assembly dependency from lower level ones, prune lower level assembly space to confirm with the fitness of higher level functional operations towards overall maintenance. This being the same mechanism section 1.3.4 unpacked to resolve the Downward Causation critique, that is, higher-level constraint families do not act on the lower level by introducing new causal forces, but by setting the parameter values and boundary conditions under which the lower-level constraint family operates, conditioning the lower-level distribution $P(X)$ on the higher-level state Y to yield $P(X|Y)$. Levin (2023, p. 12) provides numerous examples of such top-down control, such as “animals artificially made with extra copies of their entire genomes, which produces huge cells, still exhibit normal size” through overall informed cellular adjustments that maintain organ-level functionality (illustrating once more functionality dependent integrity as the final judge of viability), or — in kidney tubule formation — when cells are made abnormally large through polyploidy, “*different underlying molecular pathway is triggered (cytoskeletal bending, vs. cell–cell communication) when the cells are made so large that the only way to build the structure is for one single cell to wrap around itself, leaving a topological lumen*” (p. 12), illustrating cellular not only cellular problem solving but also conformity to overall higher level needs. Another illustration which taps into its complexities with Assembly Indexes, despite eukaryotes being a bottleneck in the tree assembly history pathway, once a tree assembly is eventuated, with it comes functional system-level operations that ensure overall integrity. Lower level cellular state-space is then pruned against the fitness of the role it plays in the functional operations that optimise towards higher-level integrity and free-energy minimisation. Sieve tube cells, for example, are typical plant cells which start by possessing a nucleus, cytoplasm, vacuole, and other organelles, yet — as they differentiate — their nucleus, tonoplast (vacuolar membrane), ribosomes, and Golgi apparatus degenerate and are lost, so as to give way to

⁹⁹ A higher-level distribution $P(X|Y)$ reweighting the lower-level distribution $P(X)$

¹⁰⁰ Which lead to higher level systems in assembly

an open, low-resistance pathway for the movement of sugars within the plant¹⁰¹. This illustrates – from the bottom-up perspective in assembly – regardless of how trees stand as much higher assembly indexes dependent on the history of eukaryotes, the nesting of cellular homeostatic drives towards integrity and information gain will direct same-level collaboration up to the phase transition in which beyond-cellular system level goals (that promote general and distributed cellular homeostasis) emerge, and, with these (system level goals), collective cellular operations in function of those goals' efficacy, which in turn shape lower level state space to assemblies better suited to said goals (such as transporting sugars).

All this to come what Levin's (2023) conceptualised as "*multiscale competency architecture*", providing a framework for understanding how individual cognitive drives (operations to persist in integrity at non-equilibria indexes) persist across assembly levels, at modular stable operations' levels which themselves seek to promote autonomous integrity, often by tapping into the advantages of being connected to broader networks (in interoperational harmony, akin to how a single physical configuration in section 1.2.1 can sit as a high-betweenness bottleneck in the assembly graph of many distinct higher-index objects). Thus, in a way highly compatible with Assembly Theory (albeit at the biology level), Levin showcases that "*what evolution is really searching [for] is not the enormous space of all possible local rules* [(see, how it also thinks of assemblies as a search object? Basal Cognition and Assembly Theory share deep affinities and synergies to be explored)], *but instead the space of behavior-shaping signals* [read in this thesis' framework above as constraint-shaping configurations whose path-reweighting channels future construction] *by which cells hack each other's functionality*" (p. 3).

The persistence of individual cognitive drives across assembly levels, still endowed with dynamic drives towards integrity at each modality, reflects what Levin (2023) identifies as a fundamental feature of biological organisation. This capacity ensures that even as cells become integrated into higher-level assemblies, they retain their fundamental nature as cognitive agents navigating problem spaces in service of integrity maintenance (how they become part of, and their relationship with, a broader higher-level organism is going to be explored at section 2.4).

¹⁰¹ "In agreement with the function of sieve elements (SEs) in long-distance transport which favors an unobstructed tube with minimal flow resistance, differentiating SEs undergo partial autolysis (...) lose its nucleus, vacuole, and Golgi apparatus, as it connects to the existing sieve-tube system" (Knoblauch et al., 2018)

2.3 From collaborator to generator

Although it would be easy to make a *straw-man* out of Minsky's *hard-wired and fixed procedure* view of agential cognition — as *passé* and ripe for reform¹⁰² — it is critical to pin down his ideas as pivotal for the current “agential/agentive” era gaining tracking in biology and artificial intelligence. Pre-Minsky, cognitive processes were regarded as unified or centrally organised. Structuralist¹⁰³ and functionalist¹⁰⁴ frameworks posited holistic mental operations — throughout most of the 20th century, cognitivism retained an implicit centralisation bias, from its inception as a Turing Machine like process to Fodor's domain-specific modules under a “central systems” coordinator¹⁰⁵, or Newell & Simon's serial symbolic architectures¹⁰⁶ — even across the tenuous disciplinary divide into Psychology, supposedly tripartite models such as the Freudian one presupposed an unified agency. Minsky's agentive insight — though naturally situated within its paradigms' limitations — by proto-regarding cognition as a form of collective intelligence, made possible the revolutionary leap into scale-free cognition.

In “Society of Mind”, Minsky (1986), leverage of the concept of *society* (and bureaucracy) to illustrate an intricate environment of mechanistic and reductionist *agents*¹⁰⁷ — seen as processes¹⁰⁸ — whose different sets of simple tasks, when articulated, give rise to the complex operations of the mind¹⁰⁹. This underscores not only the possibility of agencies with relative *simplicity* and *narrowness* (in the types operational-restraints — subsection 2.1 — they perform) at a lower-level towards higher-level cognitive processes¹¹⁰, but, as importantly the role of their articulation. By stating that “*each mental agent by itself can only do some simple thing that needs no mind or thought at all*” (Minsky, 1986), these *agents* assume the role of simple operational processes. One need not be left wondering if these basic fundamental units of cognition could be performed by non-biological entities, as the author asserts that “*there is not the slightest reason to doubt that brains are anything other than machines with enormous numbers of parts that work in perfect accord with physical laws. As far as anyone can tell, our minds are merely complex processes*” (Minsky, 1986). It follows that the substrate — the *machinery* on which these agents' tasks run and operate — of the *agent* is secondary to the operations it performs.

Minsky's insight that cognition emerges from agent distributed coordination aligns with the trajectory being drawn in this thesis. The bioelectric communication networks

¹⁰² These critiques are directly approached later on the last Minsky's subsection of this thesis

¹⁰³ “The strong version was due to the inspiration of Wundt for whom the individual mind was a synthetic unity constituted by real processes of “psychic causality.” Mental contents were the attributes of this system and could be used to study its constitution. In the weaker version of some of Wundt's pupils, like Titchener, the reference to real causal processes was dropped, but the individual mind remained the only conceivable locus for mental contents. In fact mental contents retained their status as that which constituted individual minds, although the nature of this constitution was interpreted positivistically rather than in the terms of Wundt's principle of “psychic causality.” (Danziger, 1990, p. 72)

¹⁰⁴ “I propose the hypothesis that pain, or the state of being in pain, is a functional state of a whole organism.” (Putnam, 1967, p. 226); “No organism capable of feeling pain possesses a decomposition into parts which separately possess Descriptions of the kind referred to in (2) [2 = Every organism capable of feeling pain possesses at least one Description of a certain kind (i.e., being capable of feeling pain is possessing an appropriate kind of Functional Organization)]” (Putnam, 1967, p. 226)

¹⁰⁵ “we have seen that much of what is typical of the input systems is more or less directly a product of their informational encapsulation. By contrast, I'll claim that central systems are, in important respects, unencapsulated, and that it is primarily for this reason that they are not plausibly viewed as modular.” (Fodor, 1983, p. 103)

¹⁰⁶ By introducing the idea of a General Problem Solver (GPS) — In “Human Problem Solving” (1972) — Newell and Simon emphasise centralised control (a single system applying rules step-by-step) and symbolic methods like means-ends analysis (breaking problems into subgoals), arguing intelligence arises from logical, top-down manipulation of symbols, not distributed or embodied processes.

¹⁰⁷ “each mind is made of many smaller processes. These we'll call agents.” (Minsky, 1986)

¹⁰⁸ “GRASPING agents want to keep hold of the cup. BALANCING agents want to keep the tea from spilling out. THIRST agents want you to drink the tea. MOVING agents want to get the cup to your lips.” (Minsky, 1986)

¹⁰⁹ “When we join these agents in societies [connected in] special ways (...) leads to (...) intelligence.” (Minsky, 1986)

¹¹⁰ “whenever we find that an agent has to do anything complicated, we'll replace it with a subsociety of agents that do simpler things” (Minsky, 1986)

explored in section 2.2 demonstrate this kind of distributed, agent-based organisation that Minsky theorised years prior. When cells coordinate through gap junctions and voltage gradients to achieve morphogenetic goals, they implement the same organisational principles that Minsky identified in cognitive systems, that is, simple agents (individual cells) performing specialised tasks (maintaining voltage states, responding to gradients, navigating coordinate specific problem-spaces) that collectively produce complex outcomes (coordinated tissue formation) irreducible to individual agent capabilities. Individual cellular agents maintain their specialised constraint-satisfaction procedures (homeostatic drives) while participating in distributed networks that minimise free energy at higher system levels, the kind of bureaucratic organisation Minsky explored.

Agent as a term framed via this lens — denominating an entity which functions as a cross-level unit of cognition by performing a given task — is useful as a unit which can both be further decomposed while also being capable of consideration as a unit of output. To provide an example, an agent can be an individual within a team or a team within a firm, depending on the granularity of the analysis. It is especially useful because it is a term free from biocentric biases which expresses well both centralised and distributed entities. Minsky's approach, by emphasising the importance of relationships between agents — their organisation and architecture — streamlines and harmonises individual sub-system tasks' towards higher system level processes.

The agent concept gains additional depth when understood through Assembly Theory's framework of historical constraint accumulation. Having understood cellular differentiation representing an assembly bottleneck where cells commit to specialised constraint patterns (like sieve tube cells losing their nuclei for transport efficiency), Minsky's agentic collaboration types are very likely to represent cognitive bottlenecks, where specific constraint-satisfaction procedures at scale become stabilised and reusable. Think of cognitive modalities as collaboration types, cross individual constraint pattern that can prove successful across multiple assembly contexts, becoming a modular solution that can be redeployed without re-deriving its constituent operations. As per section 1.2.1, a cognitive modality is a high-betweenness substructure in the directed acyclic graph of minimum-path realisations across many distinct higher-order cognitive assemblies. And complementing with section 1.3.3, the same substructure also sits at a stable post-singularity phase of its constraint family, which is what makes it reusable in the first place. A clear example (bowtie architectures) is explored at section 2.5.

Taking a closer look at Minsky's book: *"Whenever we speak about a mind, we're speaking of the processes that carry our brains from state to state. And this is what makes minds appear to us so separate from their physical embodiments: it is because concerns about minds are really concerns with relationships between states — and this has virtually nothing to do with the natures of the states themselves. This is why we can understand how a society of agents like Builder will work without knowing the physical constitution of its agents: what happens depends only on how each agent changes its state in response to its previous state and those of the other agents that connect to it. Other than that, it does not matter in the least what are the individual agents' colors, sizes, shapes, or any other properties that we could sense. So naturally minds seem detached from physical existence. It doesn't matter what agents are; it only matters what they do—and what they are connected to"* (Minsky, 1986) — Expresses the author, showcasing the primacy of relationships in cognition, reducing the mind to a flow of processes through states instantiated between an *agential bureaucracy*, while stressing the principle of substrate

independency; *“It’s much the same in other realms; we need to be able to describe things at many levels of detail. We all belong to families or companies, and sometimes we can think of each group as “nothing but” a network of agreements and relationships. But when we need a larger view, as when thinking about the politics of an entire country, we cannot think effectively without regarding entire families or companies as though they were single objects in a different realm. The same applies to how we think about our minds. Even if you knew all the details of each little agent in your brain, your higher-level processes would still need coarser summaries”* (Minsky, 1986) — illustrating the author’s awareness of levels of analysis that stems from an approach based on emergent phenomena, and the principle of it being scale-free¹¹¹, given that layers of complexity can be analysed from focusing at different levels of a given emergent phenomena. Minsky’s “coarser summaries” being in line with the coarse-grainings formalised in 1.1.4, maps that take fine-grained states to macrostate labels and replace a distribution over a larger alphabet with a distribution over a smaller one, with the higher-level probability content given by pushforward. It also showcases the comfort of the author in venturing into higher level agencies at the societal scale, such as “families” or “companies” as agents’ units.

To gather the full grasp of implications derived from this last quote, it is worth to linger on *Emergence* in Minsky’s line of reasoning. Given that each agent performs a single narrow task¹¹², any given system-level behaviour (i.e. involving the collective coordinated efforts of various agents) cannot be deduced from an agent’s individual properties. The author underscores this idea of rudimentary contributions towards sophisticated intelligence with the statement *“you can build a mind from many little parts, each mindless by itself”*, and a view of *“intelligence as a combination of simpler things”* (Minsky, 1986). As per Chialvo’s definition (Chialvo, 2010), *“Emergence refers to the unexpected collective spatiotemporal patterns exhibited by large complex systems. In this context, ‘unexpected’ shows our inability (mathematical and otherwise) to derive such emergent patterns from the equations describing the dynamics of the individual parts of the system”*. The term “unexpected” refers to the difficulty (or impossibility) of predicting the macroscopic properties of the system based on its microscopic properties¹¹³. The implication is that the emergent property cannot be simply derived by summing or directly analysing the contributions of individual components, given that it exhibits properties that are more than the sum of its parts. As per section 1.3, “unexpected” can be understood as emergent macrostates appearing at non-analyticities of the constraint family, where small changes in the parameters governing the joint distribution produce qualitatively new macrostate structure. The same fact is registered from the information-bottleneck side as bifurcations of the optimal representation. What looked like the impossibility of “summing” agent contributions is, within this framework, the fact that the joint distribution sits at a post-singularity phase whose macrostate alphabet is not the Cartesian product of agent alphabets. The emergence of novel cognitive properties through agent interaction mirrors the emergence of novel morphological properties through cellular coordination. Like bioelectric networks enable cells to construct forms (like two-headed planaria) that transcend individual cellular capabilities, agent networks enable the construction of mental processes that transcend individual agent capabilities.

¹¹¹ *“In neural systems, power-law distribution of avalanche sizes and durations (lifetimes) have been observed both in experiments [1,4,8] and computational models [17–19].”* (Faqeeh, Osat, Radicchi, & Gleeson, 2019); *“Power laws have an additional attraction. They are scale free. This means that they look the same at all magnifications.”* (Schulman, 2022, p. 78) — my emphasis

¹¹² In Minsky’s conception, although this is complemented and reformed throughout this thesis

¹¹³ Macrostates at a singularity-node (1.3.3) are non-analytic in the parameters of the underlying constraint family, so no smooth aggregation of microstate-level descriptions reaches them, which is exactly what makes their emergence “unexpected” in Chialvo’s sense.

Given Minsky's symbolic approach to agents, deriving emergence's scale-free nature in "*bureaucracies of agents*" can be done linearly. The author provides this relation implicitly, by stating the following chain of ideas: "*atoms in the brain are subject to the same all-inclusive physical laws that govern every other form of matter.*" (...) "*can we also explain what our brains actually do entirely in terms of those same basic principles? The answer is no*" (...) "*even if we understood how each of our billions of brain cells work separately, this would not tell us how the brain works as an agency*"; "*principles that operate at higher levels of organization*" (...) "*must add to our knowledge about lower levels*" [my emphasis] (Minsky, 1986). by stressing that the individual components of the mind are *mere matter*, subject to the same physical laws as anything else, it is not only reiterated the substrate-independence of cognition, but the subjacent idea that within the properties of *physics* lies potential cognition. This idea is key when scaling across *agent* levels: the properties of a type of task within an inside-the-skull agent contain sections of the recipe for a cognition type that can be maintained in its function at higher level agencies. The author uses across chapters the example of a child building a tower out of wooden blocks. *Builder* is an agent, within the child's mind, that is composed by further — narrower — agencies (such as find, see, move...) ¹¹⁴. As per the author, "*If you were to watch Builder work, from the outside, with no idea of how it works inside, you'd have the impression that it knows how to build towers. But if you could see Builder from the inside, you'd surely find no knowledge there. You would see nothing more than a few switches, arranged in various ways to turn each other on and off.*" (Minsky, 1986). These types of "*switches*" can be performed by entities that "*from the outside*" would be perceived as an agent that performs a section partaking to higher cognition at agglomerated scale, across other agencies, while sitting at higher level: e.g., a country's urban planning housing building agency, composed of *finding* (with groups of analysts who specialise in identifying and source materials and lots), moving (with human actuators enriched with machinery to move said materials) and so on, performing the narrow tasks of agencies underlying cognition, however at a cross-scale, cross-domain level.

Thus, Minsky's *agents* — as a collective of articulated sub-system basic processes — are capable not only of bypassing the need for biology and reducing the notion of mind ¹¹⁵ to concrete sets of functional processes, but also open the door to cognition at broader scales. From the framework described on the previous chapter, Minsky's *agents* can be conceived *informationally* as task dependent modular sub-system nodes — each restricting and channeling state space with narrow collaborative dependencies — handling subsets of tasks so as to reduce local uncertainty. Each agent's state distribution carries some local negentropy (departure from uniformity over its own alphabet), and the mutual information between agents — together, the system's total correlation — is what distinguishes a Minsky bureaucracy from a population of independent agents. And given that Minsky conceives such entities as giving origin to higher-level cognitive processes, these contribute to higher-level homeostasis drives via an overall lower-level collaboration, functionally promoting the maintenance of the higher-level system, lowering entropy at the system level scale.

Leveraging the social analogy which gives rise to *the Society of Mind*'s title, Minsky refers to these relational dynamics as "*Agents in a Bureaucracy*". The relational, information

¹¹⁴ "*building a tower is too complicated a job for any single, simple agent*" (...) "*child's mind contains a host of smaller minds. Mental agents*" (Minsky, 1986)

¹¹⁵ "*as far as I'm concerned, the so-called problem of body and mind does not hold any mystery: Minds are simply what brains do. Whenever we speak about a mind, we're speaking of the processes that carry our brains from state to state.*" (Minsky, 1986)

centric notion of *bureaucracy*¹¹⁶ (as a network of interdependencies from which collaboration, within a distributed process, articulates towards supra-individual goals) is paramount in explaining the contribution of agencies to overall cognition¹¹⁷, and will be further explored in section 2.5, but in brief, domain-independent agencies derive their utility, meaning and explainability from the context, the protocol, the network they are situated at, within the functionality of the goal being pursued, and the cognitive system in which they are inscribed.

These principles laid by Minsky resonate with agencies throughout levels generating cognitive outcomes across different scales — from a single mind to a societal institution — with agents situated at any scale, further decomposable into sub-level agencies, unified by local operations that constrict the alphabet of *by default* states to a reduced set of states that minimise free energy, lowering uncertainty towards the achievement of stable, higher-order goals.

This bridges the gap between individually held operational constraints actively maintained towards integrity and the emergence of collective and coordinated collaborations between lower level non-equilibrium assemblies which are generative of higher level cognitive processes. Demonstrating how constraint propagation and active inference principles scale across levels of organisation beyond equilibrium through agent-based architectures.

2.4 A Kaleidoscope of Selves

Assembly bottlenecks create reusable constraint patterns that accelerate future organisational complexity, and coordinated agent networks follow the same principles to create *competency* bottlenecks, that is, stable coordination patterns that enable navigation of specific problem spaces. When cellular agents coordinate through bioelectric networks toward morphogenetic goals, they're not just implementing predetermined programs but actively constructing new navigational capabilities in morphological problem space. In long term assembly routes, systems gain the ability to traverse possibly landscapes that were previously inaccessible to individual agents operating alone.

From an informational first principles standpoint, after granting each agent its own valence and local homeostatic goals — and thereby embedding cognition across scales, as shown above — the question of *how to define a system's boundaries* remains unresolved. When framed through the framework above, traditional heuristics (such as “this membrane outlines the cell” or “the skin outlines the body”) prove inadequate when considering a formalisation anchored in assembly, information, free-energy minimisation and the post-singularity macrostate compression developed in section 1.3.. This boundary problem becomes acute because agent networks generate capabilities that transcend individual agent boundaries. The Active Inference mechanisms explored in section 2.1 operate not just within individual agents but across agent networks, creating distributed prediction-error minimization systems that span multiple scales simultaneously. It is clear that rigid morphological delineations underrepresented the fluid, nested processes through which constraints, memory, and problem-solving capacities are dynamically combined across levels. Think of the way in which slime mold (*Physarum polycephalum*) solves puzzles, given

¹¹⁶ “In “Society of Mind”, Minsky (1986), leverage of the concept of society (and bureaucracy) to illustrate an intricate environment of mechanistic and reductionist agents — seen as processes — whose different sets of simple tasks, when articulated, give rise to the complex operations of the mind.”

¹¹⁷ “It is not enough to explain only what each separate agent does. We must also understand how those parts are interrelated” (Minsky, 1986)

that, as it grows, it leaves behind an externalised slime trail that it can sense and avoid, effectively offloading information to the environment about where it has been (Reid, 2023), or how Single-celled bacteria can behave as a coordinated group through quorum sensing, collectively switching gene expression towards new group level states (Ostovar & Boedicker, 2024)¹¹⁸. These examples illustrate how constraint propagation — the fundamental mechanism identified in our informational framework — operates across multiple channels simultaneously. The slime mold doesn't just navigate physical space; it navigates an information landscape where past exploration creates constraints that guide future navigation. Similarly, bacterial quorum sensing creates informational constraints that channel collective behavior toward group-beneficial outcomes. Both cases demonstrate how individual agents participate in constraint networks that generate collective competencies irreducible to individual capabilities.

The limitations of traditional boundary concepts become evident when we examine how cognitive systems actually operate in relation to their environments. The shift from individual agents to coordinated networks necessitates a corresponding shift in how we conceptualize system-environment relationships. Traditionally, from Artificial Intelligence to Cognitive Science and Psychology, the relationship of a system with its environment has been made resorting to the concept of a Task Environment. This approach — as highlighted by Allen Newell and Herbert A. Simon¹¹⁹ — stressed that the Environment is not abstractable from the agent's tasks, a notion that implicitly demarcates an agent from its borders while simultaneously looping the *outside* with an *internal* generative model. James J. Gibson perfectly illustrates the essence of this approach with the concept of *Affordances*¹²⁰, referring to the triggers for action that the environment *offers* to an agent. Given that such environmental possibilities are relational, they depend on both the properties of the environment, the capabilities and characteristics of the agent, and the priming of its goals¹²¹. Yet, although this approach embraces the principle that the definition and structure of an environment are intrinsically connected to the tasks and objectives of the agent, it has fallen short of informationally defining the agent in relation to the field of putative attainable *goals* it can pursue. This is specially limitative given that, through the concept of affordances, the environment is funneled via task oriented avenues primarily contextualised in the space of 3D spatial navigation, object manipulation, and other such immediate spatial interactions.

The task environment framework's limitations become particularly apparent when applied to the agent networks described by Minsky. Individual agents within a bureaucratic architecture don't just respond to external affordances—they create affordances for each other through their coordinative relationships. Borrowing from Minsky's (1986) simplified examples, the "Builder" agent's capacity to construct towers emerges not from its relationship to external wooden blocks, but from its coordination with "Find," "See," and

¹¹⁸ Higher level offloading of cognition to the environment and collective problem solving at the human level will be explored later at section 3.1.

¹¹⁹ "The term task environment, as we shall use it, refers to an environment coupled with a goal, problem, or task (...) It is the task that defines a point of view about an environment, and that, in fact, allows an environment to be delimited. Also (...) we shall often distinguish the two aspects of the theory of problem solving as (1) demands of the task environment and (2) psychology of the subject." (Newell and Simon 1972, p. 55). That is, there is no abstract, agent-independent environment in a given agent's generative model, there is no *single objective* representation of the environment separate from how an agent uses it.

¹²⁰ "affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill" - they exist in the environment, "independent of the individual's ability to perceive" but always measured with respect to the animal's capabilities, that is, "It implies the complementarity of the animal and the environment." (Gibson, 1979, p. 127)

¹²¹ Ellis & Tucker (2010) examined how object perception automatically primes associated actions, attesting the idea that seeing an object does not just activate its recognition, rather it primes potential actions begged by the agent's goals. And while anchored to the goal, the perception of the environment is determined by the characteristics of the system, as showcased by Linkenauger et al. (2009) by reachability perception changing when arm length is modified using tools.

"Move" agents. The environment that matters for cognitive function is increasingly the internal environment of agent relationships over the external physical environment. From a first-principles perspective, understanding non-equilibrium systems' relationship with the environment — while holding onto the nested, multi-scale nature of assembly space — implies managing a fluid relationship between *co-created internal* and *goal-relative external* environments. This conceptualisation begs for a broader scope of multidimensional spaces in which to contemplate assembly dynamics — from information gain to free-energy minimization — beyond sensorimotor navigation, which aren't contemplated by the *task environment* approach. Shifting from a *task environment* perspective to a *problem-space* framework provides such a basis for mapping a broader universe of informational interactions, in which scale-free agents operate, beyond conventional navigation and irrespective of their level. This shift represents a fundamental reconceptualisation of what constitutes *environment* for cognitive systems. Beyond a fixed external context within which agents operate, the problem-space framework treats *environment* as a multidimensional possibility landscape that is actively constructed and modified through agent interactions. The environment becomes less a container and more a field of potential constraint configurations that agents can explore, modify, and navigate.

For example, the problem-space framework reveals how the bioelectric communication networks¹²² enable cellular agents to collectively navigate possibility landscapes that transcend individual cellular capabilities, they're collectively constructing navigational competencies in multidimensional problem spaces. As an illustration, cell communication and cytoskeletal adaptation in morphological problem solving is an informational problem space: in "*Picasso tadpoles*" (tadpoles "*in which in which all the organ positions have been scrambled*"), "*the organs (...) move in novel paths and different distances, as needed, to achieve a normal frog craniofacial morphology*" (Levin, 2023, p. 1879). And while this may be at first sight perceived as 3D sensorimotor navigation space, given that it is a problem space composed of same level systems, it would be as much a navigational problem as a company of settlers deciding how to cross an unforeseen river, i.e. while involving movement and navigation at scale, it consists at the individual level of a communication and consensus making problem-space. Another good example of a non navigational problem space is transcriptional space. "*In planarian flatworms exposed to the non-specific potassium channel blocker barium, neurons die of cytotoxicity due to excessive de-polarization, resulting in rapid degradation of the head. Remarkably, when kept in barium, they soon regenerate new, barium-resistant, heads. They do this by changing the transcriptional levels of just a handful of genes. Because planaria never experience barium in the wild, there has been no selection pressure to evolve a response to barium poisoning. This indicates that when confronted by a new physiological stressor, these cells manage to navigate a path in a very high-dimensional transcriptional space*" (Mathews et al., 2023). So demonstrating how collective agent networks generate navigational competencies in abstract possibility spaces. The "Picasso tadpoles" case shows cellular agents collectively solving a constraint satisfaction problem where the goal state (normal morphology) remains fixed while the pathway to achieving that state must be dynamically reconstructed. This represents a form of collective problem-solving that transcends individual cellular capabilities. While the planarian example demonstrates navigation in a possibility space that has no evolutionary precedent. The system's ability to find viable pathways through transcriptional space under novel constraints suggests that the collective agent network

¹²² explored in section 2.2

possesses genuine creative problem-solving capabilities, not merely implementing predetermined programs. This aligns with our earlier observation about cells freed from bioelectric coordination (Xenobots, Anthrobots) spontaneously generating novel behaviors, while hinting again at the multi scale competency architecture.

A *problem-space* approach allows for the analysis of a fluid “*from part of the environment to part of the agent*” dynamic succession of lower to higher order assemblies. It does so by regarding fields of agentic *capability* in a multivariable landscape, shifting the question from “what is the agent’s task environment?” to “what *dimensions*¹²³ compose the problem space the nested system must solve to retain integrity?”. This approach emphasises that an agent’s borders should be analysed in function of the *dependencies’ field* from which attainable homeostatic states derive. It shifts focus from “immediate physical affordances localised in the external environment framed via higher-level goals” (task environment approach) to a comprehensive representation of all potential states — both realized and latent — that systems (composed by agencies endowed with valence at every scale) can influence. This reconceptualisation also reveals how the assembly bottlenecks identified in Chapter 1 scale at *non-equilibrium operations as constraints* levels as problem-space bottlenecks. When successful constraint configurations become reusable across multiple contexts they generate new competencies that enable exploration of previously inaccessible cognitive and therefore assembly *spaces*. To be useful, however, this *problem-space* approach requires a framework that can capture both the dynamic multi-scale competency of agent networks and their capacity for solving beyond 3D navigation, in multivariable problem-spaces¹²⁴. Michael Levin’s Cognitive Light Cone provides this framework, offering a dynamic boundary concept that scales across levels¹²⁵.

Micheal Levin (2019) conceptualises such problem-space approach — to the computational boundary of scale-free agential systems — in his *Cognitive Light Cone* framework, representing the spatio-temporal scope of an agent’s goal-directed activity (across scales)¹²⁶, encompassing both the present state of the agent and the set of potential future states it can influence¹²⁷. In the same manner that a physical light cone defines the limits of events observable from a given point in space-time¹²⁸, the Cognitive Light Cone defines the horizon of a system’s cognitive engagement with the world in terms of what variables it can measure, influence, and model¹²⁹, that is, the system’s ability to sense, model, predict and act within a specific variable-spatio-temporal domain. Contrary to the internal/external divide in the Task Environment model, the *Cognitive Light Cone* extends far beyond the immediate physical surroundings of the agent to include the internal states,

¹²³ What are the variables the system can leverage to retain integrity

¹²⁴ As seen above, such as a metabolic or transcriptional problem spaces

¹²⁵ While maintaining connection to the informational substrate found at chapter 1.

¹²⁶ “the scale and types of goals that a system can pursue defines (determines) the boundaries and content of the putative “agent.” On this view, what defines a coherent, unified Self out of its constituent components and the surrounding environment is the set of parts that operate toward reaching specific goal states.” (Levin, 2019, p. 9)

¹²⁷ “The cognitive boundary of an Individual (...) is the most distant (in time and space) set of events that this system can measure and attempts to regulate in its goal-directed activity. It is a surface indicating what things this system can possibly care about (conversely, it defines preferences as the spatio-temporal domain of states that serve as inputs and outputs to the system). In advanced agents, it is also the boundary of the self-model.” (Levin, 2019, p. 10); “borders of the temporal and spatial events of which a given system is capable of measuring and acting map out a “cognitive light cone” (Levin, 2019, p. 19)

¹²⁸ In relativity theory, a light cone (also called a null cone) is the set of all possible light signals emanating from (or converging to) a given event in spacetime. In a spacetime diagram, this forms a double cone (future and past) delimiting what events can be causally reached by, or can influence, the event via signals (Curiel, 2019)

¹²⁹ “Each living system has a delimited “area of concern” — a region of space-time, with the organism at its center, within which its cognitive apparatus functions to take measurements and act. (...) size and shape of this cognitive boundary defines the sophistication of the agent and determines the scale of its goal directedness. (...) The shape of boundary defines each agent’s “cognitive light cone” — anything outside this region is mentally inaccessible to that system (...) edges of a given Agent’s goal space define a sort of “computational light cone” — the boundaries beyond which its cognitive system cannot operate” (Levin, 2019, p. 11 - 12)

memories, historicity and predictive models that shape its actions, making an agent's boundary not a fixed spatial or anatomical limit but a dynamic surface defined by the agent's competencies in navigating overlapping¹³⁰ fields of problem spaces. For example, a cell, with its relatively narrow cognitive abilities, operates within a small cognitive light cone, sensing only local stimuli (such as temperature, presence vs absence of certain chemicals, mechanical and electric sensing) and responding to immediate environmental triggers, in contrast, a human — with its more sophisticated memory, planning, and predictive capabilities — has a vastly larger cognitive light cone, able to process temporally extended and spatially distant information, such as planning for the future or reflecting on past experiences. However the human's cognitive light cone has dependencies on the cell's cognitive light cone, with a cell's individual goals being influenced by how the human chooses to deploy its cognitive goals¹³¹. As per the framework above, the cone's boundary thus represents the limit of the system's capacity for prediction-error minimisation. When agent networks coordinate effectively, they collectively *expand* the system's cognitive light cone by integrating their individual predictive capabilities into a larger, more sophisticated generative model of multidimensional possibility space, while this *expansion* occurs through the same constraint propagation mechanisms, that is, individual agents contributing constraint satisfaction procedures to a collective constraint network. The bioelectric networks coordinating cellular morphogenesis represent this kind of collective light cone expansion, where individual cells contribute their local sensing and response capabilities to a collective navigation system that can solve complex morphological problems.

Following a problem-space approach, the *Cognitive Light Cone* of a system is not restricted to traditional 3D spatial navigation variables, nor fixed at a certain scale, rather, systems — whether cellular, individual, swam or collective — can operate a diverse range of problem spaces across nested *inter-agent environments* that seek to promote localised or system wide goals, in function of subsystem or system maintenance and integrity. Given the diversity of restraints and variables across the scales of assembly, the fields of problems being navigated go from metabolic to transcriptional¹³² at lower levels¹⁰⁶, to traditional higher level navigation as encompassed by the *Task Environment* framework and (as this thesis argues) governance problem spaces at higher societal level. Each of these *problem spaces* are composed of different properties and sets of variables, however, agential operations *navigating* across problem spaces share the primal homeostatic function of integrity and uncertainty reduction. To provide an example beyond those present at the footnotes, think of how organisms — such as *E. coli* and *Saccharomyces cerevisiae* — navigate a metabolic problem space consisting of optimising fluctuating nutrient availability — where they solve for efficiently metabolising different carbon sources by proactively predicting upcoming metabolic demands. When exposed to a primary sugar, these microorganisms initiate a metabolic assembly pathway that primes the cellular machinery for the second expected sugar, thereby dynamically and proactively leveraging predictions towards reducing lag time when the second sugar appears, smoothing metabolic transitions and maximizing energy extraction (Mitchell et al., 2009). This is a free-energy minimising operation at a non sensorimotor problem space — with a set of variables taking place at a broader system level

¹³⁰ "Individuals can overlap – the same biophysical system can support a number of coexisting, coupled Selves with different cognitive borders." (Levin, 2019, p. 12)

¹³¹ Like forgetting to take the anaemia pills, sorry cells

¹³² "Single cells are very good at managing their morphology, behavior, and physiology as needed for survival, altering their motility, and metabolism in response to, and proactively in, changing environmental conditions. They succeed in exploiting their microenvironment toward optimal reproduction by selecting among numerous possible choices of gene expression patterns and behaviors." (Levin, 2019, p. 5 - underscore emphasis by me)

environment, in which *actions* are defined informationally. The implications of navigating such information problem-space — if further research showcases similar behaviour in multicellular organisms' cells — reverberate across scales of systemic integrity. These problem spaces are defined as fields given that higher level assemblies have dependencies on the role of lower level processes, sharing overarching assembly goals, albeit in incommensurable scales encompassing different sets of variables and time scales. Think of a national level *bureaucracy* attributing to the health ministry *agency* the responsibility of promoting healthy states across its individual human constituents, meaning making sure that these humans' promote healthy states towards the general nurturing of organ and cell function and integrity — that perform its homeostatic functions towards human maintenance — all the while simultaneously honoring the overall governance goal of having a healthy population base with which to run its constitution and thus perpetuate the national state, each cross-scale agency operating in a field of accessible problem-spaces. The systems' operations prompting homeostasis and minimising free-energy — i.e cognitive — are not constrained by physical boundaries but are instead shaped by the informational problem-space they can generate, influence and navigate. This governmental example reveals how cognitive light cones can nest and overlap across scales. The health ministry's cognitive light cone encompasses both the immediate operational variables of healthcare delivery and the broader temporal variables of population health trends. Individual human citizens operate within their own cognitive light cones while simultaneously serving as components within the ministry's expanded light cone. Each higher-level cone being a coarse-graining of the cones beneath it, with the macrostate alphabet at each level being the equivalence classes the higher-level inquiry treats as indistinguishable. This nested structure demonstrates how higher-order cognitive systems emerge through the coordination of lower-order systems without eliminating the autonomy of constituent agents.

This information centric¹³³, cross-goal and scale-free¹³⁴ *system boundary* has profound implications for understanding *selfhood* and *agency* as dynamic¹³⁵, nested¹³⁶ and shaped by the *system's* ability to navigate overlapping problem space fields to maintain integrity. All the while, by defining *agent and system* boundaries in purely informational terms (constraints, uncertainty reduction and generative model), their formalisation becomes grounded in the same first principles established in Chapter 1, ensuring a monist integrated approach¹³⁷. By anchoring in first principles of assembly and information, it becomes clear that previous rigid, morphological, delineations underrepresented the fluid, cross synergy, nested processes through which constraints, memory, and problem-solving capacities are dynamically combined across levels¹³⁸.

The Cognitive Light Cone framework establishes the dynamic scope of system competencies, but it requires a complementary statistical framework to understand how these competencies maintain coherence across scale transitions. The challenge is

¹³³ "what defines this Self is the boundary of information being able to pass between the subunits" (Levin, 2019, p. 8); "definition of an Individual based on its information-processing structure" (Levin, 2019, p. 9)

¹³⁴ "somatic cells did not lose their behavioral plasticity and computational capabilities in becoming part of metazoan swarms (bodies): they scaled them to enable pursuit of larger goals" (Levin, 2019, p. 5); The more a cell is connected to other cells in networks, the more processing capacity and the bigger the horizon of what the compound individual can potentially sense, remember, and store (...). Merging into communicating networks also enables more noise tolerance and robustness (Levin, 2019, p. 17)

¹³⁵ "These boundaries are malleable, and can shift at different time scales, to sizes limited by what the underlying hardware supports." (Levin, 2019, p. 21)

¹³⁶ "systems and subsystems will have different, overlapping, and not always aligned goals. Thus, another key implication is that systems house multiple levels of coexisting coherent Selves." (Levin, 2019)

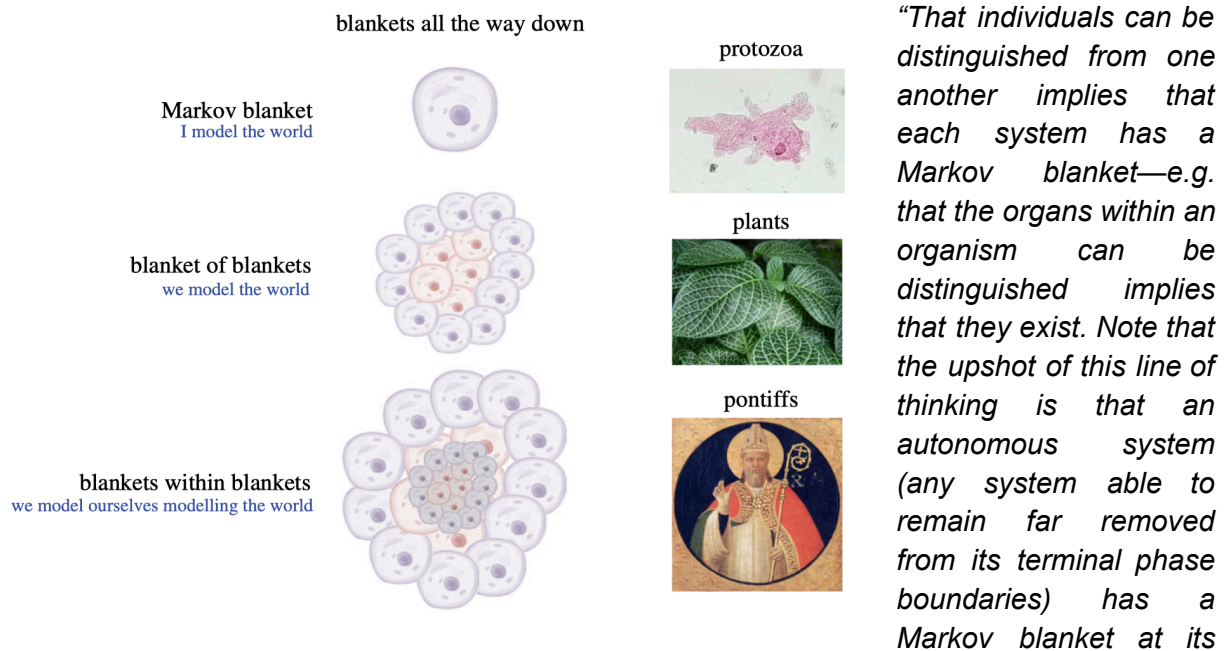
¹³⁷ Rather than relying on legacy morphological labels, while sustaining the logical incorporation — as developed in the following sections — of seemingly intangible informational concepts relegated to culture and language

¹³⁸ More on the individual human delegation of cognitive load to higher level structures on the next chapter

explaining how systems with expanded cognitive light cones maintain their integrity while operating across multiple problem dimensions simultaneously. The solution lies in recognising that cognitive boundaries are fundamentally statistical. When agent networks coordinate to expand their collective cognitive light cone, they create statistical patterns of interaction that define system identity through conditional dependencies. This insight bridges the earlier discussion of bioelectric coordination (or, more broadly, the functional role of *Cognitive Glue*, as explained throughout this thesis) with the broader question of system boundaries by revealing how coordination creates statistical signatures that persist across scale transitions. As the conventional biocentric *organism* gives way to a wider range and spectrum of systems, so does the notion of *individuality* shift. Whereas — stemming from the bias tied to viewing and interpreting the world from a higher level human perspective — an organism is traditionally viewed as an individual entity, a more objective observation would readily note that the individuality of a system is better described by patterns of conditional independence and correlation of subsystems. When considering the array of systems compatible with the framework afforded by Assembly Theory, the idea of “individuality” transcends mere spatial or morphological delineation — like a cellular membrane or a body outline — and instead emerges as a statistical construct defined by functional relationships and independencies. Individual systems emerge as stable patterns in the constraint propagation networks established in Chapter 1. Systems are thus defined by the statistical signatures of their constraint satisfaction procedures. Equivalently, in the framing of section 1.3.3, individuality is the macrostate that consolidates at a stable post-singularity phase of the constraint family governing the system's components, a phase that registers at the higher level as “a system having formed”. In essence, individuality in such systems can be characterised by the presence of a *Markov Blanket*¹³⁹, a probabilistic partition that demarcates the internal states of a system from its external environment, with a set of *blanket* states filtering all exchanges between inside and outside. Markov Blankets were introduced as a tool in Bayesian networks, but as our understanding of information merged with biology and cognition it has since been extended under the *Free Energy Principle* and *Active Inference* frameworks — its significance lying in how it frames the system's autonomy and its capacity to actively maintain internal organization amid environmental perturbations — by defining which states are internal and which are external, thereby carving out an informational boundary that identifies the system as a self-coherent entity. When systems minimise variational free energy they're actively constructing and maintaining their Markov Blanket boundaries through their prediction-action cycles. For instance, a given mammalian organism's internal metabolic states and neural dynamics are shielded behind its sensory and active states, ensuring that any influence from the environment must pass through these blanket states — thus preserving a stable sense of self and continuity over time. Karl Friston and colleagues have highlighted that such living systems — by continually inferring and adapting to their environments — effectively build generative models of the world through these Markov Blankets, minimising variational free energy. The authors of “*The Markov Blankets of Life: autonomy, active inference and the free energy principle*” illustrate this particularly well with two figures and accompanying text:

¹³⁹ (Kirchhoff et al., 2018)

Figure 4: Kirchhoff et al., 2018 - 1 (in paper figure 3)



superordinate level composed of Markov blankets at its supraordinate level. The self-evidencing dynamics of autonomous organization can therefore be cast as exhibiting two different yet complementary tendencies: an integrative tendency of a multiplicity of Markov blankets to self-organize into a coherent self-evidencing whole, and a self-assertive tendency to preserve individual autonomy. This is the basis for the claim that autonomous systems are made up of Markov blankets of Markov blankets. Central to the idea of an ensemble Markov blanket is that the statistical form and subsequent partitioning rule governing Markov blankets allow for the formation of Markov blankets at larger and larger scales (of cells, of organs, of individuals, of local environments). The reason for this is that the organization of Markov blankets occurs recursively at larger and larger scales, recapitulating the statistical form of Markov blankets at smaller microscopic scales" (Kirchhoff et al., 2018, p 8)

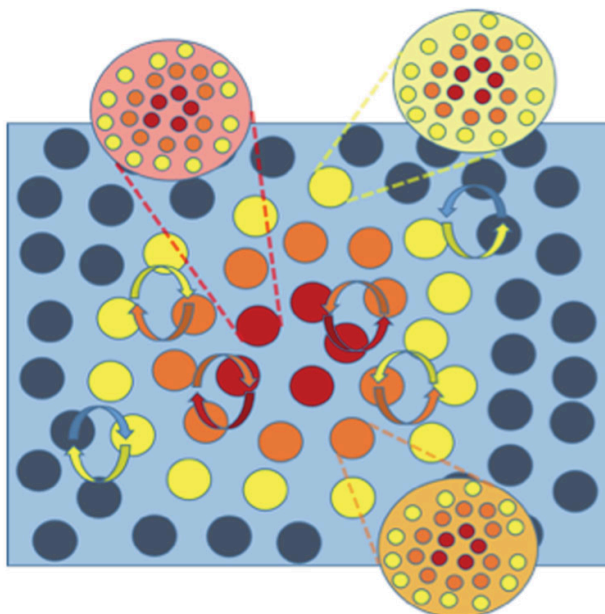


Figure 5: Kirchhoff et al., 2018 - 2 (in paper figure 4)

"a system constituted by a multiplicity of nested Markov blankets at the scale of microscopic dynamics, and a larger or bigger Markov blanket at the macroscopic scale of collective dynamics. It thus becomes possible to distinguish between internal and external states only by appeal to the presence of a third set of states; namely, the

Markov blanket. This means that the assembly of Markov blankets can be understood to occur in a nested and hierarchical fashion, where a Markov blanket and its internal states at the macroscopic scale consist of smaller Markov blankets and their internal states at microscopic scales of systemic operations. Crucially, the conservation of Markov blankets (of Markov blankets) at every hierarchical scale enables the dynamics of the states at one scale to enslave the (states of) Markov blankets at the scale below, thereby ensuring that the organization as a whole is involved in the minimization of variational free energy. It is thus only when the properties of the collective dynamics feed back into the scale below, forming a free energy-minimizing system at the scale of the whole system, that it is possible to talk meaningfully of ensemble Markov blankets—blankets whose self-evidencing dynamics result in an overall self-sustaining organization.” (Kirchhoff et al., 2018, p 8)

The recursive structure of nested blankets provides the statistical foundation for understanding how Minsky's agent bureaucracies maintain coherence across levels, revealing also how the multi-scale competency architecture explored in section 2.2 maintains statistical coherence across scale transitions. Individual competencies don't disappear when agents coordinate, rather becoming components of higher-order competency architectures that preserve individual capabilities while enabling collective functions. The recursive nesting of Markov blankets formally instantiates the recursive coarse-graining grammar of 1.1.4, in the sense that each higher-level blanket is the equivalence class produced when the lower level's joint distribution is grouped into macrostates, and each level's probability content follows by pushforward through the chain of coarse-grainings that connects it to the base. Importantly, the statistical individuality defined by a Markov Blanket need not correspond strictly to traditional physical or anatomical boundaries. Instead, individuals can be hierarchically composed of “Markov Blankets of Markov Blankets”. In biology as we know it — and as illustrated by the figures above — this is illustrated by cells, tissues, organs, and entire organisms each manifest as nested layers of statistical autonomy, in a broader framework as supported by Assembly Theory however, its Markov Principles may remain in architectures as we are yet to know or conceptualise. This hierarchical nesting further supports the multi-scale point above, conveying this aspect also to individuality, allowing it to emerge at multiple levels of a system's configuration — *all the way down* to molecular assemblies and *all the way up* to beyond human ones (such as Governance Cognitive System's, as argued by this thesis) or other configurations at scales currently out of humanity's grasp. Each level being an adaptive agent in its own right, maintaining its own *Markov Blanket*. And further sustaining the point above that, blanket-defined identities also support transcending the rigid life/nonlife dichotomy, with systems not conventionally considered *alive* forming Markov Blankets and processing information — adaptively reducing free energy gradients — due to statistical configurations. Thus, *individual* systems emerge as stable statistical structures under selection and adaptive inference — capable of resisting the eroding forces of thermodynamic equilibrium — by continuously modulating interactions with their environment, rendering individuality not a static attribute, but a dynamically maintained statistical relationship, reflective of how a system manages its own operational closure while actively engaging with its environment.

This *individuality boundary* can nest and overlap at multiple scales. Cells within an organism, for example, manifest their own Markov Blankets — regulating ionic concentrations, gene expression, and local metabolism — while also participating in tissues or organs that themselves maintain more expansive Markov Blankets. As seen, this nesting

disrupts the notion of *individuality* as a monolithic entity while also allowing a subsectioning of individuality into further agents. The human microbiome, for example, exemplifies this multi-agential nature, with bacteria, archaea, and fungi inhabiting skin, gut, and mucosal surfaces, each population maintaining its own Markov Blanket while integrating (partly or wholly) into the host's physiological and cognitive processes — given that host–microbiome crosstalk influences metabolism (Devaraj et al., 2013), immune function, even cognition (Chen et al., 2021), challenging simplistic notions of a single, physically circumscribed individual (Rees et al., 2018) — thus challenging, from a purely anatomical vantage point, the traditional notion of where the *self* ends and the *microbe* begins, while from a statistical vantage point, these sub-communities exhibit autonomous boundaries while remaining coupled to the larger host system. The microbiome example is particularly revealing because it demonstrates how statistical boundaries can transcend species barriers while maintaining functional coherence, illustrating both individual dynamic drive and the multi-channel constraint structure of higher-level consolidation discussed in section 1.2. The microbial communities don't lose their individual autonomy when they integrate into crucial constraint operations within system hosts' functions, they maintain their own drives of individual constraint satisfaction procedures while contributing to higher-order constraint networks. What emerges is a multi-level mosaic of individuated adaptive subsystems — each minimising its own free energy within the broader context of the *host's* Markov Blanket, in webs of conditional independence. This field of dependency drives makes it possible to understand problem-spaces as interwoven networks of information processing — thus the importance I've given to both downwards and bottom-up causation throughout this thesis — as it opens our eyes to, as Levin (2019, p. 9) put it, “*what defines a coherent, unified Self out of its constituent components and the surrounding environment is the set of parts that operate toward reaching specific goal states*”. In fact, even a *body's* outline itself is only one expression (among others) of selfhood, and in many contexts — such as in insect colonies — Markov Blankets of Markov Blankets produce ensemble-level *organization* that behave as a higher-scale agencies, with its own adaptive goals. So, prior to an individual always stands a selected scale of observation.

2.5 Cognition, Governance or principles of operating integrity?

When a system's cognitive light cone expands through agent coordination — as when cellular networks navigate morphological problem spaces — what informational infrastructure enables this expansion while preserving individual agent autonomy? In striding towards this answer stands the recognition that multi-agent systems share a fundamental coordination problem. Agents operating within their local cognitive light cones must somehow align their constraint-satisfaction procedures with system-level goals they cannot directly perceive or model. A hepatocyte navigating metabolic problem space cannot directly sense predator threats, yet must respond appropriately to stress signals. This coordination gap — between local competencies and global requirements — necessitates informational infrastructure that can compress system-level states into locally interpretable signals while preserving functional coherence across scales. While section 2.3 introduced Minsky's notion of *bureaucracy* as a metaphor for understanding how simple agents coordinate toward emergent higher level cognition, this section aims to move past the mere metaphor to understand *bureaucracy* as a functionally definable *class of informational solutions* to the coordination problems inherent in multi-agent systems operating far from equilibrium.

In line with the informational framework developed in the sections above, the concept of *bureaucracy* can be freed from its legacy context — as originated in eighteenth-century France, derived from the French word "bureau" (referring to desk with drawers typical of office furniture) and the suffix "-cracy" (from the Greek word *kratos*, meaning power or rule), to express "*rule by writing desk*"¹⁴⁰ — and functionally framed from an information centric perspective. Multi-agential systems have, due to their distributed nature, shared sets of problems, such as asymmetric access to information across agents, the need for storage and retrieval of beyond-individual-level information, the necessity for organisation and categorisation of such information, and the harmonised translation and enforcement of said information into agent's state-space and output. These are addressed by a constellation of *compression*, *categorisation*, *shared reality creation* and *adaptive interpretation* solutions functionally analogue to the role of *bureaucracy* at the inter-human scale.

It is important to acknowledge that the exploration undertaken within the next paragraphs was not pursued by Minsky — who employed the idea of a *bureaucracy* more as a superficial metaphorical tool from which to explain the articulation of *agents* towards general cognition, a limiting factor of his framework, alongside others I'll be soon making explicit — whereas this thesis takes an ontological, assembly inspired, information centric and functionally defined view of *bureaucracy* as a scale-free constellation of solutions towards higher level articulation of lower level agencies (enabling system-order lower entropy and maintenance).

One solution employed across scales, in multi-agential systems, is *information compression* through bowtie architectures, in which multidimensional realities are compressed into simplified, actionable schemas in which fidelity is traded for salience to allow for efficient communication, interpretation and coordination. As per Levin (2024, p.9), "*this bowtie architecture — which forces a compression of data to a generative kernel that then has to be re-inflated and elaborated — is (...) a common feature of biochemical, bioelectrical, and biomechanical pathways*". For a concrete example, in morphology, "*Each organism does not make a copy of itself—it undergoes a massive compression, producing an egg, which then has to re-inflate during the process of self-assembly and engage with a possibly different world using possibly different (mutated) parts*" (Levin, 2024, p. 6). Bowtie compression solutions are information processing frameworks wherein complex, high-dimensional inputs are compressed into low-dimensional representations before being expanded into context-appropriate (more on interpretability later) outputs, allowing for storage efficiency, transfer (communication) flexibility and efficacy, robustness to damage, integration across scales and innovation¹⁴¹. Bowtie architectures instantiate the constraint propagation mechanisms established in Chapter 1 within a multi-agent coordination context. The compression phase reduces the alphabet of possible system states, generating information through entropy reduction (as formalized in section 1.1). The bottleneck plays the role described in section 1.3.4 as a path-reweighting locus, the configurations that pass through it inherit their probability mass not from a uniform prior over physically possible configurations, but from the realised assembly distribution that has selected for persistence and propagation. The reusability of the bowtie kernel is the assembly-graph signature of this reweighting (section 1.2.1). From the active inference perspective (section 2.1), individual

¹⁴⁰ "Bureaucracy literally means "rule by writing desk." The term was invented in eighteenth-century France, when the typical official sat next to a writing desk with drawers—a bureau. At the heart of the bureaucratic order, then, is the drawer. Bureaucracy seeks to solve the retrieval problem by dividing the world into drawers, and knowing which document goes into which drawer." (Harari, 2024)

¹⁴¹ While each claim is addressed in the paragraphs below, the section on innovation is developed in the paragraph dedicated to interpretation.

agents cannot maintain generative models of the entire system state, as such models would exceed their computational capacity and cognitive light cone boundaries. Instead, the compressed bottleneck provides a sufficient low-dimensional representation that preserves the information necessary for local agents to minimise their prediction errors relative to system-level goals. The p53 node, for instance, doesn't require individual cellular agents to model the full complexity of DNA damage patterns, metabolic states, and stress contexts, it provides a compressed signal that cells can interpret to reduce surprise relative to their homeostatic preferences (Lane & Levine, 2010). This reveals why bureaucratic solutions are not merely convenient but necessary for multi-agent systems: they emerge as inevitable consequences of the free energy principle operating across agents with bounded cognitive light cones. Systems that fail to develop effective compression-decompression architectures cannot maintain the statistical coherence (Markov blanket integrity) required for persistence at higher assembly indices.

Regarding storage efficiency, by reducing the dimensionality of stored information, systems minimise the energetic and material costs of information maintenance. For example, instead of maintaining separate, dedicated signaling cascades for every possible environmental stimulus, cells employ *hub genes*¹⁴² that serve as convergence points for multiple upstream signals. Going back to the example of the p53 tumor suppressor network, a multiplicity of different cellular stressors (DNA damage, hypoxia, nutrient deprivation, oncogene activation...) are detected by diverse sensor proteins, but instead of maintaining separate response pathways for each input type — which would require enormous metabolic resources — these signals converge on p53 (Lane & Levine, 2010). This compression reduces the dimensionality from perhaps hundreds of potential inputs to the activation state of a single regulatory node, so, as depicted in Levin's bowtie architecture diagram, p53 serves as the bowtie *bottleneck* (Levin, 2024, p. 7, figure 2, B). As per transfer (communication) flexibility and efficacy, Compression allows for information transfer across mediums that would normally block the transmission of detailed, high-dimensional representations. Levin (2024, p. 9) references that "*a very simple parameter — waves of ATP — can mediate the complex morphogenetic information being shared and processed within groups of embryos defending their anatomies against teratogenic influence*", illustrating how bowtie architectures enable coordination in biological collectives through minimal signaling bandwidths, with the complex multidimensional state of "anatomy is under threat" being compressed into ATP, then decompressed by receiving embryos into appropriate defensive morphogenetic responses. By compressing complex states into transferable signals that trigger contextual interpretation, bowtie architectures enable sophisticated coordination even when direct, detailed information transfer is impossible. What appeared in section 2.2 as coordination enabling collective navigation of morphological problem spaces can now be understood as bureaucratic information architecture: bioelectric networks implement bowtie compression-decompression cycles that allow cells to participate in system-level goals while maintaining their individual cognitive autonomy. The electrical signals that link melanophores and xanthophores in zebrafish pigmentation, or coordinate gap junction networks in tissue patterning, function as bureaucratic messages—compressed representations of system states that individual cellular agents can interpret within their local cognitive light cones.

¹⁴² Hub gene: "A gene that is highly connected to other genes in a network, and is thought to be important for the regulation of the network as a whole" — ScienceDirect. (n.d.). Hub genes. In ScienceDirect Topics. Retrieved October 5, 2025, from <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/hub-genes>

The robustness to damage conferred by bowtie architectures is exemplified in planarian flatworm regeneration, in which — despite (or perhaps because) of possessing what Levin calls "*an incredibly noisy genome*" (2024, p. 10) — planarians have remarkable regenerative capacity, resisting "*transgenesis, aging, cancer, and injury*" (2024, p. 10). When a planarian is cut into fragments, each piece can regenerate a complete organism, *because* the animal's morphological information is compressed into bioelectric patterns that persist even when physical structures are disrupted. This compression allows the system to maintain function despite significant perturbation, as "*a planarian fragment (...) obeys the bioelectric pattern memory of the original body, but only carries a small part of it when it is cut out*" (p. 6). From a multi-agent perspective, this illustrates how diverse cellular collectives (fragments which can be regarded as Markov Blanket agencies) can reconstitute coordinated functionality by interpreting *compressed* morphological instructions. So as to not require every cell to possess complete, detailed blueprints (which would be vulnerable to mutation and damage), the system distributes a *compressed representation* that remains interpretable even when substantial portions of the network are lost. This pattern of resilience through compression (although with dependencies on interpretation) showcases design principles adopted by function multi-agent systems that must maintain coordination despite component failure or information corruption. The planarian example reveals how bureaucratic architectures function as assembly bottlenecks in the section 1.2.1 sense. The bioelectric pattern memory represents a stable constraint configuration that has proven viable across evolutionary assembly space, a reusable solution to the morphological coordination problem that can be propagated forward without exhaustive re-exploration of developmental possibility space. When planarian fragments regenerate, they're not searching the full combinatorial space of possible cellular arrangements; they're following established assembly pathways encoded in the compressed bioelectric bottleneck. Successful constraint configurations become crystallized as high-betweenness substructures that channel future assembly exploration without forcing combinatorial re-derivation. *Bureaucratic solutions* thus represent a *specific class*¹⁴³ of assembly bottleneck, functionally coordinating multi-agent systems across cognitive light cone boundaries. Following that same principles that have the peptide bond serving as a molecular assembly bottleneck that countless protein structures reuse, the information compression and decompression method as a solution serve as an assembly bottleneck that diverse multi-agent assemblies can leverage. This explains the convergent evolution of bureaucratic information architectures (across scales), they represent stable solutions in assembly space that enable higher-index assemblies by solving the coordination problem.

As per the integration across scales, for example (Herman et al., 2003), the stress response system demonstrates such cross-scale integration via compression through the hypothalamic-pituitary-adrenal (HPA) axis: when an organism faces threat, multidimensional environmental information (predator presence, injury, social threat...) from various sensory modalities is compressed into a simplified neurochemical representation at the hypothalamic level. This representation is further compressed into a single hormonal signal — corticotropin-releasing hormone (CRH) — which serves as the minimal hub, the bottleneck, in the bowtie. This compressed signal then propagates downward, expanding into context-specific responses at cellular levels (elevated glucose production, immune modulation, mitochondrial activity adjustment...) and upward into emergent organizational patterns at the behavioral level (vigilance, motivation, memory formation...). This

¹⁴³ Among a universe of others potential solutions

architecture allows for efficient coordination among autonomous agents that could not possibly process the full complexity of information available at other scales (hepatocytes need not understand predator behavior to appropriately mobilise glucose, immune cells need not comprehend social hierarchies to optimally allocate resources). The compression hub enables multi-agentic coordination by transmitting only the essential, actionable information (threat salience over threat fidelity) while allowing each subsystem to interpret and elaborate this compressed signal according to its specific functional requirements.

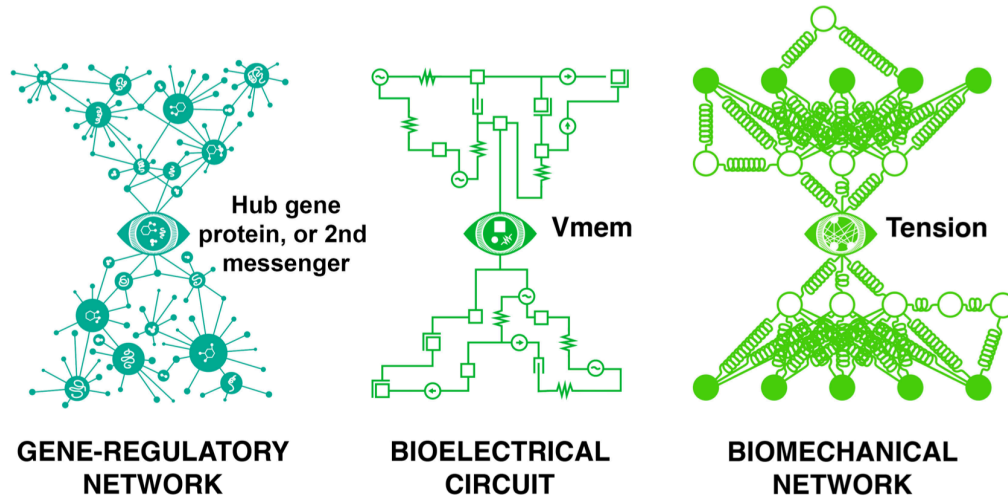


Figure 6: Bowtie Architecture and *Bowtie Bottleneck* illustration. From (Levin, 2024, p. 9), with the following description: “Schematic representations of bowtie architectures in biochemical, bioelectrical, and biomechanical circuits. For example, causal parameters such as tension or resting voltage potential across the membrane (*Vmem*), as well as specific genes, can act as information hubs that ensure the compression of signals within biological control networks.”

Once multivariable information is distilled to its *bowtie bottleneck state*, categorisation functions as a crucial bridge between *it* and meaningful action, by transforming abstract, compressed representations into structured, relational frameworks that facilitate adaptive interpretation across multiple *time* and *order* (level) scales, allowing agents distributed across these dimensions to maintain coherent deployment of information despite such heterogeneity. Categorisation emerges not merely as a passive classification system but as an active, generative process that mediates between compression-induced information loss and the requisite specificity needed for contextually appropriate action. This reconstruction doesn't rely on deterministic algorithms, as the compression process inherently eliminates precise correspondences between input and output states — as Levin (2024, p. 10) observes, “*by removing correlations, the resulting engram looks increasingly random*”— thus creating a gap that must be bridged through interpretive frameworks, therefore categorisation systems that can transform compressed near-random signals into coherent constructs are required to maintain functionality. To provide a concrete illustration — as Levin (2024, p. 4) explains — in a butterfly's metamorphosis, memories persist despite the fact that “*their hardware is strongly refactored as the creature shifts from a soft-bodied machine that lives in a 2D world and eats leaves to a hard-bodied one that flies through 3D space and drinks nectar*”. This persistence occurs through *categorisation mechanisms* that abstract their functional significance from the specific implementation details. The caterpillar's memory of “green, thin, and teardrop shaped means food” is necessarily categorised so as to not overfit

according to its perceptual features or motor responses, but according to its functional role in promoting adaptive behaviour — “*not leaves, but ‘food’*” (Levin, 2024, p. 4). This categorical abstraction enables the butterfly to utilise the same functional relationship in an entirely different sensorimotor context, while leveraging *bowtie bottleneck artifacts* as cues in new modalities of interpretation.

It is thus easy to foresee the importance of *adaptive interpretation* in the deployment of such compressed information, and its yieldings in flexibility and adaptability across scales, time and agent types, to suit present needs. As Levin (2024, p. 2) puts it, “*Biology went all in, committing to the fact that everything will change in ways that cannot be anticipated, and thus to the idea that defending static Selves is futile*”. This is evident in the case of ectopic eyes in tadpoles, where visual organs induced¹⁴⁴ at the tail connect not to the brain but to the spinal cord (or even gut), yet still enable functional vision (Levin, 2024). The tadpole thus demonstrate nested adaptive interpretability not only at the cell individual agent level — in which the cell will dynamically seek and adapt to a state space compatible with “collective eye” output, adaptively seeking solutions compatible with the non-*traditional* morphological coordinate — but also at the organ and overall system’s visual module level, adaptively developing entirely novel neural pathways. That is, interpreting compressed morphological information cues in a malleable and context dependent manner. Similarly, the differential response of tadpoles and froglets to the same bioelectric stimulus (monensin) further exemplify *context-sensitive interpretation* (Levin, 2024), where the same signal that triggers *tail growth* in tadpoles will kickstart *leg growth* in froglets, indicating that “*the same bioelectric stimulus is functionally reinterpreted based on the specific context, preserving salient, high-level information like ‘build whatever normally goes at this location’*” (Levin, 2024, p. 6). Thus, information compressed into bottlenecks is — on the other end of the bowtie — decompressed flexibly in accordance to context, across multiple scales of organisation. Such flexibility enables the necessary freedom of agents to integrate information originally so as to adaptively respond to life’s grand spectrum of original situations with resilience.

This adaptive interpretation mechanism reveals a crucial feature of bureaucratic architectures, in that they preserve agent autonomy through controlled ambiguity. The compressed bottleneck signal is intentionally under-specified, it doesn’t dictate exact implementation details but rather constrains the possibility space within which local agents exercise their problem-solving capacities. The “build whatever normally goes at this location” instruction exemplifies this. It’s a bureaucratic directive that compresses system-level anatomical goals into a form that individual cellular agents can interpret using their local generative models. From the active inference framework, this means bureaucratic bottlenecks function as priors that shape (but don’t determine) agent behavior. The CRH signal doesn’t mechanistically force specific cellular responses—it updates the priors in cellular generative models, biasing their active inference loops toward stress-response configurations. This preservation of inferential autonomy is what distinguishes bureaucratic coordination from mere mechanical coupling: agents remain free to minimize their own prediction errors, but within a probability landscape reshaped by system-level constraints.

Finally, these mechanisms of compression, categorisation, and adaptive interpretation converge in shared *reality creation*: a collective process by which distributed agents, across time and scales, establish coherent and interoperable frameworks of understanding despite having different *perspectives* (as context dependent interpretations) and asymmetric access to information. “*For example, consider a rat trained to associate*

¹⁴⁴ (Blackiston & Levin, 2013)

pressing a lever with receiving a reward. No individual cell has both experiences — the skin cells of the paws experience the lever, the cells in the gut receive the metabolic reward. This associative memory belongs to the “rat” — the collective — and to none of the cells individually” (Levin, 2024, p. 13). The shared rat level *reality* is not a static representation present by default at each individual agent (organ, cell, organelle), but an actively maintained, dynamically reinterpreted cognitive construct¹⁴⁵. As Levin explains, “*while the details of minds and bodies change, thought forms (salient cores abstracted from experiences [i.e. the compressed bottlenecks dynamically interpreted per context and level]) remain and survive because they move across architectures that are able to interpret them in new ways for sense-making as a gestalt, rather than committing to their details*” (2024, p. 7). This allows agentic systems to achieve a coherent self-identity despite continuous material turnover and environmental flux (the ever material changing of the system while maintaining the *system*). This dissolves the traditional distinction between the *cognitive machinery* and the *information* it processes, framing instead a continuum of assembly, wherein “*certain kinds of information structures actively facilitate their transformation and remapping in ever-changing cognitively excitable media*” (Levin, 2024, p. 7). This reconceptualises system’s *selfhood* as an active process of interpretation and *reality creation*, deliberately not a static entity, as Levin (2024, p. 12) puts it “*memories [a bottleneck compression subject to interpretation steering and contextualising output], especially non-local ones, actually reify the agent to whom they belong, in a sort of feedback loop in which the Self elaborates and maintains memories, which in turn reinforce its existence beyond that of its parts*”.

While the rat-level self emerges from cellular agents coordinating through bioelectric and biochemical bottlenecks despite no individual cell possessing the full experience, governance-level assemblies emerge from human agents coordinating through documentary, cultural and legal bottlenecks¹⁴⁶ despite no individual human possessing the full institutional memory or systemic perspective. The mechanisms are functionally isomorphic across scales: compression of multidimensional realities into actionable representations, categorisation that enables distributed interpretation, and adaptive deployment that preserves local autonomy while maintaining collective coherence. Human societies face the identical coordination challenges that cellular collectives solve through bioelectric networks: asymmetric information access, cognitive light cone limitations, the need for persistence across component turnover, and the requirement to maintain system integrity despite environmental perturbation.

These types of *informational* solutions showcased above are far from being exclusive to cell agents, and are rather part of a broader suite of design principles addressing cooperation and maintenance at multi-agentic systems. At the human level, bureaucracy *compresses* multidimensional realities into simplified, actionable schemas in which fidelity is traded for salience¹⁴⁷ — where diverse inputs converge into a simplified kernel¹⁴⁸ — in the form of records and registry entries, likewise enabling storage, transfer capability (communication) and cross-scale integration¹⁴⁹. Categorisation too is a fundamental

¹⁴⁵ Further enforcement and citations developing this idea are offered in the section “A Kaleidoscope of Selves”, in which the concept of Xenobot and Anthropot showcase the potential independence of cell agencies, unbound from the overall *body reality*

¹⁴⁶ The next chapter will focus on this shared, broader, *cognitive light cone* binding human level assemblies into Governance ones

¹⁴⁷ “bureaucracy too tends to sacrifice truth for order” (Harari, 2024)

¹⁴⁸ “Anyone who ever filled out an official form knows this only too well. When you fill out the form, and none of the listed options fits your circumstances, you must adapt yourself to the form, rather than the form adapting to you.” (Harari, 2024)

¹⁴⁹ Family to municipality, municipality to country, country to international market

bureaucratic operation¹⁵⁰, abstracting principles and selecting primary unifying vectors from the compressed bottlenecked information into structured, relational frameworks so as to enable coordinated action across agents, privileging distributed access to information over its integrity, tolerating granularity loss for actionability and accessibility. Such categorisations create artificial, but necessary, order. For example, the effect that "*history is separate from biology and from mathematics (...) is the intersubjective invention of academic bureaucrats*" (Harari 2024), disciplinary boundaries that enable specialisation and cooperation while also – due to artifacts of granularity loss – create artificial separations that don't reflect reality's continuity (and have *reality building* repercussions, akin to those showcased in organism systems)¹⁵¹. So as biological systems structure responses into functional interpretable *schemas* per bottleneck's categorisation¹⁵², human level bureaucracies devise taxonomies¹⁵³ to locate and action responses through bowtie bottlenecks (documents) efficiently. A great example is provided by Harari (2024), describing how the 1854 London cholera outbreak was resolved by John Snow's categorisation of cases by location — a bureaucratic act that transformed raw data into actionable insights. Less heroic or dramatic examples are yielded daily by bureaucrats around the world deploying bottlenecked information to such scripted societal behaviour as providing subsidies or deploying security forces.

Flexible responses to compressed signals, based on context, are equally enabled by the adaptive interpretation of rules, forms, and procedures according to context, with laws and regulations being actuated based on circumstance (the role of the judiciary, as opposed to the regulator branch, is to interpret and casuistically deploy legal and bureaucratic informational bottlenecks so that they are suitably deployed into the richness of life). As seen above, where biological systems reinterpret the same signals differently depending on developmental stage or environment, adaptive bureaucratic interpretation allows bowtie bottleneck artefacts to be deployed flexibly across diverse contexts, adapting to novel situations, maintaining system flexibility¹⁵⁴. The worldbuilding implications of bureaucracy are perhaps the easiest to intuitively grasp when framed through the implication that "*to own a field (...) mean that it is written on some clay tablet, bamboo strip, piece of paper, or silicon chip that you own that field*" (Harari, 2024)¹⁵⁵.

Property rights, citizenship, legal status, borders and such¹⁵⁶ are built and sustained constructions of governance cognitive systems' through shared bureaucratic bottleneck compression, categorisation and interpretation, yielding the actively maintained *selfhood* of state, municipality, institution and so on, in a feedback loop akin to the biological one

¹⁵⁰ "Bureaucracy seeks to solve the retrieval problem by dividing the world into drawers, and knowing which document goes into which drawer" (Harari, 2024)

¹⁵¹ "Scientists and legislators have categorized species according to the threat of extinction they face, on a scale ranging from 'least concern' through 'vulnerable' and 'endangered' to 'extinct.' Defining a particular population of animals as an 'endangered species' is an intersubjective human convention, but it can have far-reaching consequences" (Harari, 2024).

¹⁵² As illustrated in the paragraph on categorisation above

¹⁵³ In the form of drawers, shelves, digital folders...

¹⁵⁴ Personal story: at an examination during my Law Bachelor, in a case of *Unjust Enrichment* due to privileged access to leaked information about an eminent change in *zoning status* to a property, I've applied Water Distribution legislation for the restitution. Water distribution rights, with origins in Roman Law (Dig. 8.3.0. De servitutibus praediorum rusticorum), have long codified the *res communes* principle of access to water (from there I've extrapolated the principle of those closer to the *leak* source having *obligations* to those in the stream further from it). Such principle has been reinterpreted and adapted from Medieval to Modern Legal Systems (Convention on the Law of the Non-navigational Uses of International Watercourses; Adopted by the General Assembly of the United Nations on 21 May 1997), showcasing the casuistic and context dependent interpretative power of agents dealing with bureaucratic bowtie bottlenecks.

¹⁵⁵ The informational bowtie bottleneck in the form of the "clay tablet, bamboo strip, piece of paper, or silicon chip" that carry the bureaucratic information isn't enough to sustain a common worldbuilding framework coordinating action across individual human agents towards a societal governance cognitive system, further "collective cognitive glues" are discussed throughout the Part II of this thesis.

¹⁵⁶ "If your neighbors have been grazing their sheep for years on a piece of land, and none of them ever said that you own it, but you can somehow produce an official document that says it is yours, you have a good chance of enforcing your claim." (Harari, 2024).

described above, in which the *self* — as an articulated consortium of agents — elaborates and maintains bowtie bottlenecks, which in turn reinforce its existence beyond that of its parts (people come and go, states remain). That is, that which remains is the credence, value, common interpretation and deployment of bowtie bottleneck artifacts which yield common consensus, articulation and a shared *self* world building at the higher societal level¹⁵⁷. An interesting artefact of this reliance on informational compressions — such bowtie bottlenecks from corticotropin-releasing hormone (CRH) to clay tablets — in world building, is that the compressed bottleneck ends up not solely *representing* reality but it *becomes reality* instead: when a person suffers a panic attack derived from no objectively real threat, cell agents will still respond to CRH with the whole suite of stress responses that are employed when real threats are present; when an individual contracts no debt but the registry faults and places debt in their name, they will carry the same responses of having debt despite having no debt in the *real world*.

The Romanian Jewish census of 1938 provides a haunting example of how bowtie bottlenecks shared reality creation can have life-or-death — or, in a sense, autoimmune responses — consequences: "On January 22, 1938, the government issued a law ordering all Jews in Romania to provide documented proof that they were born in Romanian territory and were entitled to Romanian citizenship. Jews who failed to provide proof would lose their citizenship, along with all rights to residence and employment" (Harari, 2024). (This resulted in ~367,000 Romanian Jews losing their citizenship, many of whom then were murdered during the Holocaust). Closer to my example sits Australia's Robodebt scheme (Royal Commission into the Robodebt Scheme, 2023). These governance failures parallel biological coordination failures at a mechanistic level. The Romanian census and Robodebt scheme represent bureaucratic compression faults—situations where the categorization step (citizen/non-citizen, debtor/non-debtor) eliminated context necessary for accurate decompression, creating interpretive frameworks that mapped individuals into action scripts incompatible with their actual circumstances. At the organism level, organ rejection too — for example — can be framed as a lower level *world building* failure, a fault of interpretability due to information loss during compression, in which immune cells through low-bandwidth receptors (e.g TCRs, PRRs), interpret bowtie bottleneck artifacts (e.g donor MHC proteins, DAMPs) in a "self/non-self" or "danger/safe" classification, due to both the evolutionary compression algorithm (e.g., TCR-MHC binding avidity, cytokine thresholds) lacking in granularity to encode "beneficial foreignness" or "new system", and due to the cellular agents' impossibility to interpret beyond the horizon of information compressed to their functionality. Although not made from an informational, bureaucratic or even bowtie model conceptualisation, Medzhitov and Janeway's (2002) description of the innate immune system's three-fold recognition strategy of "microbial nonself", "missing self" or "induced or altered self" informed my example. Their work demonstrates how pattern recognition receptors (PRRs) decipher molecular signatures to either induce or inhibit immune responses. MHC proteins function as critical "molecular passwords of self-identity", where their absence triggers NK cell activity against "missing self". The authors note that while this discrimination strategy effectively distinguishes normal self from nonself, it "is not resistant to fraud," suggesting inherent limitations in its interpretative capacity. The "fuzzy decision-making mechanisms" they describe, where susceptibility of target cells depends on "the balance of the ligands for activating and inhibitory receptors" reflects an evolutionary

¹⁵⁷ "National myths legitimize the tax records, while the tax records help transform aspirational stories into concrete schools and hospitals" (Harari, 2024).

compromise between sensitivity and specificity that lacks granularity for recognising beneficial foreign tissue

Having a transplanted organ be rejected is then a Kafkaesque tragedy from an information processing perspective, similar to other bureaucratic horror stories of compression and decompression faults in documentation. Take UK's Windrush Scandal, when citizenship was compressed discarding context in a manner which didn't accommodate for Windrush-generation migrants status — which was tied to colonial-era oral agreements — and treated archival gaps as fraud, stripping legal residents of healthcare, jobs, housing and threatening them with deportation (Townsend, 2018)

Throughout scales, the need for distributed information to yield coordinated action leads to sacrificing realism for order — enacting stress responses in organisms when no threat is present or executing a debt collecting script when no funds were allocated — yet it enables vectors of information to be maintained and shared across scales and time, generating bowtie bottlenecks that outlast individual components so as to create responses that transcend agents' lifespans towards higher level system's integrity, a general dependency of having lower level systems performing actions without holding complete information, nested in higher level wide integrity and goals. Across scales, what matters is not a perfect representation of reality (a 1:1 scale of the world's variables is impossible anyway, without becoming the environment itself¹⁵⁸) but functional informational architecture adequate for problem-solving across agents' nested goals.

From a first-principles, assembly theory and active inference perspective, this suite of informational architecture solutions — i.e. bureaucracies — arise from the properties of non-equilibrium systems composed by multiple agencies, with their fundamental constraints on information processing and nested homeostatic needs being the *origin* of the adaptive protocols that constitute functional solutions. By offering a common operative framing, it likewise diminishes surprise¹⁵⁹ while lowering the degrees of freedom of each individual agent to the spectrum of action scripts afforded by the bowtie bottlenecks (within suitable interpretations). From an *outside the script* perspective, such bowtie bottleneck's affordances constitute an imposition of constraints from *all possible responses* to a small set of compatible actions — whether through chemical bonds, bioelectricity, or organisational policy sustained by documentation — playing the role described in section 1.3 as configurations at which the relevant constraint family consolidates, with the path-reweighting of section 1.3.4 selecting which lower-level configurations remain compatible with higher-level assembly integrity. Through path-dependency dynamics, higher-level assemblies are shaped by the contingencies of such bureaucratic information architectures, with multi-agentic organisms with unfit information distribution solutions being nonviable. Thus, such informational bureaucratic architectures aren't an imposed human invention but an instantiation of scale-free, entropy-reducing solutions for multi-agent coordination that is an informationally viable architecture for coordinating towards higher level system goals. Current governance solutions overlap with this viability. Is it necessary the only viable architecture? Possibly not.

The role of the bureaucracy overall appears then to be one of protocol and articulation between individual agential tasks towards emergent system level outcomes.

¹⁵⁸ "A one-to-one map may look like the ultimate representation of reality, but tellingly it is no longer a representation at all; it is the reality" (Harari, 2024)

¹⁵⁹ This hypothesis would explain why bureaucracies typically favor predictability over optimality — being fundamentally engaged in uncertainty reduction rather than absolute efficiency — thus favouring surprise minimisation and integrity in the face of environmental complexity over erratic yet optimal gains in efficiency that could compromise overall systemic maintenance.

From an information processing perspective — if each agent's output is treated as derived from an alphabet of potential agential states — then the bureaucratic configuration performs the role of a constraint articulation channel, filtering the fit of potential sub-system individual states at scale and sequence against a given system's level goal. It functions as a protocol in the sense that — by selection, reweighting, via uncertainty reduction in outcome or whatever other origin from which the bureaucracy protocol might originate — it selects the admissible sub-system's alphabet of states conducive to the system-level task being pursued (thus overall system integrity), however non-linear and susceptible to dynamic interpretation. In total-correlation terms (section 1.1.2), an effective bureaucracy is one whose joint distribution over agent states sits at a stable post-singularity phase, a configuration where the components are mutually constrained enough that the joint alphabet is much smaller than the Cartesian product of the agent alphabets, but not so collapsed that the agents lose the local autonomy that adaptive interpretation requires. This is compatible with the existence of individual agents whose role is to check or provide feedback about the performance of other agencies and their adherence to the bureaucratic protocol, while still being inscribed in it (which would require bureaucracies to be self-referential): a notion that highly resembles how Karl Friston derives Metacognition (meta-awareness and attentional control) from Free Energy minimisation, adapted here into Minsky's lexicon as a loop of agents whose role is to check other agency's status while being themselves checked by further higher checker agencies in a bureaucracy that enables the higher-order system to achieve cognitive opacity and deliberate attentional control (Sandved-Smith et al., 2021). This recursive, self-referential protocol enabling layered feedback highlights a key bureaucratic characteristic — agencies in a bureaucracy may serve a dual function, both as an operational unit and as a checkpoint for adjacent agents. That is, of bureaucracy as protocol pruning and constricting individual subsystem agent's alphabet of potential states — via its own internal agents — towards system-level coherence and output. Whereas each agent can be viewed as an aggregator, manipulator or restrictor of state spaces towards a reduced set of desired outcomes as opposed to default *natural* ones; a bureaucracy of agents is a self-referential protocol enacted by the agents that run it, to prune the state space of *overall individual agents*, against the fit of system level goals, so as to achieve it in emergent scale.

This self-referential structure deserves elaboration, as it connects bureaucratic architecture to the hierarchical inference mechanisms that characterize sophisticated cognitive systems. In Friston's framework, metacognition emerges when systems develop hierarchical generative models where higher levels model the reliability and performance of lower levels. A bureaucracy implements this architecture: oversight agents maintain generative models not of the external environment but of other agents' constraint-satisfaction performance, generating predictions about whether lower-level agents are successfully minimizing free energy relative to system goals. Consider the immune system's regulatory T cells (Tregs), which suppress other immune responses (Sakaguchi et al., 2008). Tregs function as metacognitive agents within the immune bureaucracy—they model the activity of effector T cells and intervene when those cells' responses threaten system integrity (autoimmunity). This is bureaucratic self-reference: agents (Tregs) monitoring other agents (effector T cells), with both sets operating within the same constraint propagation network, both ultimately serving system-level homeostatic goals. Similarly, in governance bureaucracies, auditing agencies, judicial review processes, and oversight committees (among others) serve as metacognitive agents. They maintain generative models of other

bureaucratic agents' performance, predicting whether those agents are correctly compressing inputs and appropriately decompressing outputs. When prediction errors arise — when lower-level bureaucratic agents deviate from expected performance — metacognitive agents intervene to correct trajectories, just as Tregs suppress wayward immune responses. This reveals bureaucracy as inherently hierarchical in inferential architecture. The system maintains predictions about itself, monitors its own performance, and corrects its own errors through nested layers of oversight agents, each layer modeling the reliability of the layer below, each layer subject to monitoring by the layer above. This recursive structure is what enables bureaucracies to maintain coherence across scaling transitions while adapting to novel challenges. The metacognitive layers can update their models of lower-level reliability as circumstances change, maintaining system flexibility without sacrificing coordination.

This characterisation of agents — as autonomous homeostatic systems that optionally participate in bureaucratic coordination — reconciles the apparent tension between individual autonomy and collective coherence that has threaded through Chapters 1 and 2. Agents maintain their individual cognitive drives (section 2.2), their local free energy minimisation loops (section 2.1), and their distinct Markov blanket boundaries (section 2.4), while simultaneously contributing to higher-level constraint networks through bureaucratic channels. The bureaucracy doesn't eliminate individual agency but rather provides an informational substrate through which individual agents can coordinate their constraint-satisfaction procedures toward goals beyond their individual cognitive light cones. This also points to the evolutionary trajectory from single-celled organisms to complex multicellular systems to social collectives, highlighting each transition as the development of increasingly sophisticated bureaucratic architectures (from bioelectric networks to linguistic-legal systems) that enable coordination across increasingly heterogeneous agents with increasingly disparate cognitive light cones, without requiring the elimination of individual agent autonomy. Each such transition is a singularity-node in the constraint family that governs multi-agent coordination, a configuration at which the joint distribution consolidates into a stable post-transition phase whose macrostate alphabet (organism, society, governance assembly) is not visible in any agent-level description prior to the transition. The bioelectric networks that coordinate cellular collectives represent early-stage bureaucracies, the legal-documentary systems that coordinate human collectives represent mature-stage bureaucracies, but the underlying informational mechanics —compression through categorizable bottlenecks, adaptive interpretation preserving local autonomy, shared reality creation through distributed constraint satisfaction — remain structurally invariant across scales. The cross-scale parallels between these suites of solutions beg for an unified framework for understanding multi-agentic problem solving at scale¹⁶⁰, with integrated research deploying yieldings across legacy fields. This functional decomposition of *bureaucracy*¹⁶¹ was a vector informing me on the utility of functionalism, which considerations close Part I of this thesis.

This framing of Cognition certainly survives *Occam's razor* better, as it solves various problems: first, it streamlines and illustrates the path of the smaller narrower subsystems *agents* in their sophistication to eventually enable bureaucracy building, by grounding their

¹⁶⁰ represent solutions to the fundamental challenge of maintaining coherent multi-agent systems in changing environments through the strategic processing, storage, and reinterpretation of information, in which "*intelligence = a publicly observable (third-person perspective) degree of competency at reaching the same goal by different means as necessary, with respect to a problem space and goals identified by some observer (which can also be the system itself)*" (Levin, 2024)

¹⁶¹ Whose guiding principles were laid out in section 1.1, 1.2 and 1.3, i.e. the substrate, mechanism and threshold by which the multi-agent coordination problems addressed in this section become tractable.

existence on historic stand-alone roles independent of the *bureaucracy* (thus breaking an implicit chicken-egg situation in Minsky's formulation); second, it points towards answers to nature's recurrent solution of nested system, subsystem, sub-subsystem architectures; and third, it holds explanatory power to such phenomena as *cancerous behaviour* – in which a subsystem's drive towards self homeostatic maintenance overdrives the broader system wide integrity¹⁶² – or *competition* between sub-system agents¹⁶³ (within the same system – which, even in non-pathological outcomes, showcases how even “greedy” agents can expand their measurement boundaries, effectively linking up with neighbors to form a bigger self for access to more information and resources, composing self-interested Markov Blankets^{164,165}).

2.6 Nested Interests: Alignment Across Dependent Scales

Understanding this universal coordination architecture requires distinguishing between the two complementary aspects of mechanistic constraints — that establish system-level coherence (*hard protocols*) — and the adaptive processes through which constituent agents navigate constrained possibility spaces (dynamic drives).

Let us consider the biological examples already explored above. In cellular systems, hard protocols manifest as the biochemical and bioelectric constraints that structure permissible cellular behaviors. The voltage patterns maintained by gap junction networks (section 2.2) function as hard protocols, establishing the bioelectric landscape within which individual cells operate, delimiting which cellular states remain compatible with tissue-level morphological goals. For example, when a melanophore in zebrafish skin encounters a xanthophore and its membrane depolarises, triggering migration away from contact, this represents a hard protocol, that is, a mechanistic constraint encoded in ion channel configurations and receptor sensitivities, that structures the cell's behavioural repertoire¹⁶⁶. The individual melanophore doesn't *choose* whether to respond or not to depolarisation. The biochemical architecture mechanistically compels it to. However, within these constraints, the cell exhibits *dynamic drives*, from which it actively *interprets* local voltage gradients, integrates signals from multiple neighboring cells, and navigates toward configurations that minimise its individual free energy and integrity.

¹⁶² “In cancer, the scope of the coherent Self of a cell reduces from being as large as the boundary of a whole body, to that of just one cell's surface. The scope of the Self – the structure which a cell works to maintain and support – shrinks drastically” (Levin, 2019, p. 8)

¹⁶³ “Since its discovery, cell competition has been implicated in a range of physiological processes, including cancer, aging, immunology, and development” (Gogna et al., 2015, p. 698)

¹⁶⁴ “Greed, at the single-cell level, for information (infotaxis) drives cooperativity, as each unit expands its measurement boundary (communication with neighbors) and thus inevitably becomes part of a bigger self with bigger set points serving as homeostatic attractors.” (Levin, 2019, p. 20); “The advent of multicellularity has been proposed to result from the drive to minimize surprise: a cell need only to surround itself with its progeny, in order to ensure a much more predictable milieu (the least surprising object in the world is a copy of yourself) (...) The surrounding body becomes an informational shield (Markov blanket) for the stem cell; adding epithelial and mesenchymal cell layers enables a kind of neural-like network in which each layer provides an increased level of abstraction from environmental stimuli” (Levin, 2019, p. 15)

¹⁶⁵ I have the impression that this paragraph is a significant contribution to the philosophy of mind and our definition of life

¹⁶⁶ The p53 tumor suppressor network provides another instructive example. As developed in section 2.5, p53 functions as a bowtie bottleneck compressing diverse cellular stressors (DNA damage, hypoxia, nutrient deprivation, oncogene activation) into a single regulatory hub. The hard protocol consists of the gene regulatory architecture, that is, the specific transcription factor binding sites, the phosphorylation cascades that activate p53, the downstream genes under p53 control, and the biochemical thresholds determining whether cells repair, arrest, or undergo apoptosis. These constitute mechanistic constraints that delimit the space of possible cellular responses to stress. Individual cells cannot arbitrarily choose how to respond to DNA damage, i.e. the p53 network architecture constrains permissible responses. However, within this constrained space, cells exhibit dynamic drives through context-dependent interpretation, given that the same level of p53 activation triggers different outcomes depending on cell type, metabolic state, and concurrent signals. A cell receiving identical DNA damage signals may execute different response programs because their local generative models — shaped by epigenetic state, available metabolic resources, and developmental history — lead them to different regions within the p53-constrained possibility space.

The stress response system (section 2.5) exemplifies this architecture at the organism scale through dual pathways. The sympathetic-adrenomedullary (SAM) system triggers immediate fight-flight-freeze responses via direct neural activation and catecholamine release, while the hypothalamic-pituitary-adrenal (HPA) axis orchestrates sustained responses through its hormonal cascade—CRH→ACTH→cortisol (Ulrich-Lai & Herman, 2009). This neuroendocrine architecture establishes the structural framework of threat-detecting neural circuits, parallel signaling cascades and tissue-specific receptor distributions that compress environmental information into biochemical signals propagated throughout the organism. These constitute mechanistic constraints that structure organism-level stress responses. An organism, when encountering a predator, cannot arbitrarily choose whether to activate the HPA axis or not, the neural architecture will trigger as above. Yet within this constrained architecture, individual organs and tissues exhibit dynamic drives through adaptive, situational and idiosyncratic interpretation.

System level cognition — across cellular to organism levels¹⁶⁷ — is therefore achieved from a continuous and dynamic interplay between two fundamental aspects: *hard protocols* and *dynamic drives*, as autonomous agents with individual homeostatic goals, when coordinating toward collective outcomes, face two emergent requirements. First, the need for mechanisms that constrain the joint possibility space to configurations compatible with system-level integrity, and second, the need for flexibility that preserves agent-level autonomy within those constraints. *Hard protocols* satisfy the first requirement by propagating constraints that eliminate non-viable configurations, while *dynamic drives* satisfy the second, by enabling agents to adaptively navigate the constrained space through active inference and a drive towards individual integrity.

Hard protocols therefore constitute the mechanistic constraints, architectural structures, and regulatory thresholds that delimit the joint possibility space of multi-agent systems, specifying which lower-level configurations remain compatible with higher-level assembly integrity. They establish the formal operational code of the system — the biochemical thresholds, voltage patterns, regulatory architectures, or institutional rules — that define boundaries of permissible agent states, creating the conditional dependencies that reduce system entropy below the sum of independent agent entropies. Through distributed enforcement, where agents at the same organizational level monitor and constrain each other's state-spaces, hard protocols enable peer-policing that maintains collective coherence without centralized control. They represent the stable, explicitly instantiated assembly of the informational state-space of higher-order systems, the constraint-propagation described in section 1.3.4 that compress multidimensional system requirements into signals that eliminate non-viable microstates from the accessible alphabet.

Given that higher level integrity is dependent on the integrity of its lower level agencies, the most efficient method to guarantee system perseverance and robustness is by encouraging the intrinsic homeostatic motivations of component agencies. However, maintaining their own integrity preferences shapes how they respond to system-level signals, necessitating interpretive mechanisms that reconcile individual and collective goals. *Dynamic drives* therefore constitute the adaptive interpretation processes, context-dependent decompression mechanisms, and agent-level inference procedures through which constituent agents navigate the constrained possibility space established by hard protocols, while simultaneously pursuing their own free energy minimisation and

¹⁶⁷ I am motivated to pursue this kaleidoscope to the organelle level, but was advised to chose and stick with a level of granularity compatible with a credible delivery of this thesis

integrity. They embody the active inference loops, the local homeostatic goal pursuit, contextual signal interpretation, and multiple idiosyncratic operational heuristics, through which agents decompress compressed system-level specifications into implementations. Through distributed interpretation, where agents decode shared constraint signals according to their local generative models, contextual circumstances, and intrinsic motivations, dynamic drives enable adaptive flexibility that maintain system coherence while preserving the agent-level autonomy upon which collective robustness depends. They represent the phenotypic expression layer of system architecture — the context-dependent instantiation mechanisms through which stable informational constraints (hard protocols as "genotype" from a system architecture level) yield variable behavioural outcomes (dynamic drives as "phenotype" from a system level architecture) adapted to local conditions, illustrating how identical constraint specifications decompress into different implementations depending on agent state, environmental context, and historical trajectory. This architectural relationship between stable compression (hard protocols) and variable decompression (dynamic drives) recapitulates across all scales the fundamental principle that constraint propagation preserves system identity through statistical coherence while enabling adaptive exploration within delimited possibility spaces, resolving the alignment problem by making individual goal pursuit compatible with collective persistence.

The integration of hard protocols and dynamic drives doesn't represent a binary distinction as much as a continuous spectrum. At one extreme lie purely *mechanistic*¹⁶⁸ constraints that permit minimal interpretive flexibility — electromagnetic forces holding atoms together; thermodynamic laws governing chemical reactions; physical geometries constraining molecular conformations... — while at the other extreme lie highly interpretable constraints that permit vast adaptive flexibility: bioelectric patterns specifying morphology; stress signals triggering organismal responses; constitutional provisions structuring governance (to be explored in the next Chapter). This spectrum reflects the bandwidth available for dynamic drives to operate within hard protocol constraints, from low bandwidth (mechanistic rigidity) to high bandwidth (interpretive flexibility).

This spectrum corresponds to the cognitive light cone differences explored in section 2.4. Systems composed of agents with small cognitive light cones, limited capacity to model alternatives, predict consequences or coordinate responses, require hard protocols with low interpretive bandwidth. Cellular systems largely fall in this category. Individual cells cannot model tissue-level morphological goals, predict long-term developmental outcomes, or coordinate complex collective strategies, so bioelectric networks must mechanistically constrain their behaviors through relatively rigid voltage thresholds and ion channel configurations. Systems composed of agents with large cognitive light cones — extensive capacity to model alternatives, predict consequences, and coordinate responses — can

¹⁶⁸ **Important disclosure:** The term "mechanistic" is admittedly imperfect, carrying historical baggage from mechanism-vitalism debates and potentially implying an ontological divide between physical and cognitive processes, that this thesis explicitly rejects. Even constraints described here as "mechanistic" — electromagnetic forces, thermodynamic laws — are ultimately informational in nature (as established in Chapter 1) and involve probabilistic, this not purely deterministic, dynamics. The term is retained for lack of a better alternative that captures the essential distinction: constraints operating with minimal interpretive bandwidth, where the mapping from constraint specification to instantiated outcome is relatively invariant across contexts. "Mechanistic" usefully denotes this low-bandwidth, context-insensitive end of the constraint spectrum, while acknowledging that all constraints — from atomic bonding to constitutional interpretation — operate through the same fundamental informational substrate of possibility-space compression and constraint propagation established throughout Part I.

operate with hard protocols having high interpretive bandwidth. Governance systems (that form the focus of later sections) fall in this category: individual human agents can model institutional alternatives, predict policy consequences, and coordinate collective action, enabling constitutional frameworks to specify high-level principles that agents interpret contextually, that is, not mechanistically implementing rigid behavioural scripts (although human psychology, specially at the social level, is far from lacking scripts and schemas). Notice that this change is not qualitative, or in form, but a quantitative difference on the same architectural solutions to the same multi-agency problems. The distinction is not absolute but relative. Even systems with small-light-cone agents (cells) exhibit significant interpretive flexibility within bioelectric constraints, as the ectopic eye (section 2.5) and barium-resistant regeneration examples demonstrate. Even systems with large-light-cone agents (humans in governance systems) require a great relative degree of mechanistic rigidity in their hard protocols, institutional coordination cannot be left to mere contextual interpretation without losing collective coherence and function. The balance between mechanistic constraint and interpretive flexibility constitutes a design parameter that multi-agent systems must navigate.

The subsequent Chapter examines how hard protocols and dynamic drives operate specifically at the human level, and how governance systems emerge through cultural *cognitive glue*. The functional architecture, though, remains invariant: bowtie compression-decompression cycles, adaptive interpretation preserving local autonomy, nested Markov blankets maintaining statistical coherence, active inference minimising free energy across hierarchical levels, mutual restraints promoting the positive total correlation of section 1.1.3 that is the signature of systemhood. Before turning to those specifics, however, a methodological question lingering from the preamble requires resolution. What epistemological attitude captures scale-invariant operational principles, while accommodating substrate agnostic implementations? This leads naturally to *functionalism*, as hinted above.

2.7 Functionalism across substrates

The preceding sections have shown how multi-agent systems operating beyond equilibrium require both *hard protocols* (mechanistic constraints delimiting joint possibility spaces) and *dynamic drives* (adaptive interpretation processes through which agents navigate constrained spaces). This architectural relationship — stable compression enabling variable decompression — operates across scales. Yet recognising this scale-invariant architecture is insufficient if the conceptual vocabulary employed to describe it remains anchored in scale-specific disciplinary traditions. A framework is required to capture operational principles independently of their particular instantiation — one that doesn't ask "what is this made of?" but instead "what *role* does this play?". This is the stance of *functionalism*.

Adopting a functionalist approach becomes necessary once information is selected as the *reductionist substrate* (section 1.3.4). Analysis then becomes grounded on the operations that *maintain*, *constrain*, and *update* informational states (Chapter 1), putting emphasis on *role* over material instantiation (be it biological, institutional, cultural, technological...). That is, across scales — from cellular to institutional — the principles showcased above manifest through different material substrates. Enabling both cognitive and political phenomena to be analysed from a common vantage point of informational operations. That is, as hierarchical, non-equilibrium assemblies that (i) propagate and

enforce constraints to compress large microstate alphabets into functional subsets (information as departure from maximum entropy, 1.1); (ii) accumulate those constraints across time as historical restraint-cascades and assembly bottlenecks that embed memory and enable higher-index constructions (Assembly Theory); (iii) consolidate new levels of description at non-analyticities of the constraint family (singularity-nodes), with downward causation operating through the conditioning of lower-level distributions on higher-level states; (iv) form nested Markov blankets and expand collective cognitive light-cones through multi-agent coordination; and (v) that maintain integrity by minimising free energy through active inference. This opens a¹⁶⁹ path towards understanding Governance assemblies as not distinct from those entities Cognitive Science is concerned with, unified in a common vocabulary of uncertainty reduction, constraint imposition, goal seeking and adaptive stabilisation. The bioelectric networks coordinating cellular morphogenesis (section 2.2) and the legal-documentary systems coordinating human societies (previewed in section 2.5) appear fundamentally different when viewed through their material substrates, yet *functionally* both implement the same informational architecture.

The primary obstacle to realising this unifying insight is terminological. That is, the informational framework developed in Chapters 1 provides a common logic and *syntax*, yet the legacy *grammar* of existing literature needs to be revisited and recalibrated functionally so as to leverage this newfound unity throughout the labels and categories used across disciplines. The legacy of concepts in both Cognitive and Political Science carry considerable historical, contextual and semantic burden. Terms like "memory," "archive," or "precedent" evoke a myriad of context-dependent assumptions and disciplinary priors that obscure their underlying informational dynamics. As a result, attempting to review text based literature — regardless of anchoring the review on informational principles as the ones derived above — between fields can stumble when the terms are understood as originally tied to their set of situational legacy, regardless of the operation, role or architecture underpinning their categorisation.

To truly integrate and apply the lens established above, one must adopt a principled approach to understanding categorised phenomena, one that originates from the physical and informational substrate itself, instead of uncritically importing pre-formed categories bequeathed by legacy Political or Cognitive Science. This means relying on a conceptual tool that extends (or restricts) each terminology to the role it plays. Grounding classifications in these operational processes enables an analysis that treats cognition as what non-equilibrium systems do to maintain low entropy — substrate and scale independent, informational and statistical, nested and multi-agential — instead of being concerned with what the particular biological architectures are. This liberates terminology from bio- and individual-centric legacy, providing the necessary ground to deploy and interpret concepts across the Cognitive/Governance divide. Where "memory" denotes retained constraint configurations enabling assembly reuse, "perception" denotes prediction error minimisation updating generative models, and "coordination" denotes distributed constraint satisfaction through bureaucratic compression-decompression cycles, regardless of whether instantiated through bioelectric networks or legal-documentary systems. The result is a coherent, scalable framework for understanding how varied systems maintain structural and functional integrity through the active, information-driven orchestration of constraints. Pragmatically, it means one ought to begin by asking *what does a given term* — understood as a component,

¹⁶⁹ Not the path, as many other approaches or epistemological grounds — when based off first-principles independently derived from the legacy frameworks of either Political or Cognitive science, structured ground up so as to be common to both — are also possible and potentially concurrently useful.

a type, an architecture of a system — *do in terms of managing state space possibilities*. I do not wish to abolish familiar terminology, but rather to release it from its legacy anchors to see beyond disciplinary barriers and biases. This approach was informed mainly by *Assembly Theory* and *Basal Cognition*, which, as presented throughout the chapters above, challenge foundational classic definitions such as *life* and *cognition* to assume more fundamental meanings, allowing more encompassing study-spaces founded on the pillars of first-principles, from physics to information theory.

To fully integrate knowledge across these disciplines' funnels, an exercise on functional informational taxonomy is critical to break the semantic inertia preventing labels and categories to be intelligible between them. Because no thesis functions in the vacuum, given that I am applying the term *Functionalism* — allow me first to preemptively indulge this paragraph to a meta-taxonomy of the sort alluded to above, inspecting the role of the term itself, and the reason why it is being employed here, given that it has played different roles from biology to sociology, political science to neuroscience and philosophy of mind¹⁷⁰. As per Wouters (2013, p. 482), "*Functional biologists view organisms as solutions to the problem of how to stay alive and explain their organization by investigating what is needed to stay alive and how those requirements are met*", and so — unlike the everyday understanding of functions as associated with purposes or reasons — biological *functionalism* is un-teleological, regarding functions as roles (supporting the system's resilience)¹⁷¹. So, that which the functionalist biologist seeks to explain is "*why the function bearer needs to be present, needs to have the structure it has, or needs to do what it does*" (Wouters, 2013, p. 483) as per the role it plays in homeostasis. This framework was adapted into sociology¹⁷², where society itself is regarded as a complex system composed of interrelated parts, each performing *functions* — referring to the *role*, or contribution, of specific components (such as institutions, practices, or norms) — serving a specific purpose to maintain the stability and equilibrium of the whole¹⁷³. Whereas in Political Science there is a fork on how the term is understood and employed. While on one hand, it leans into the previous understanding stemming from biology — as a framework explaining political phenomena by analyzing the *roles* (i.e. the functions) performed by political actors within a system to maintain integrity¹⁷⁴ — on the other hand, spearheaded by David Mitrany's approach, it came to signify (specially in realm of international relations) the theoretical framework of addressing international needs through non-territorially-bounded *functional* (performing specific roles) organisations, that seek international stability bypassing sovereign boundaries¹⁷⁵, with the

¹⁷⁰ "the word 'function' does not designate a single all-purpose concept of function, but in fact a plurality of (...) concepts" (Mahner & Bunge, 2001, p. 75)

¹⁷¹ "The biological function of a part or activity is its role in the organization that enables the continued existence of that organism." (Wouters, 2013, p. 483).

¹⁷² "he [Talcott Parsons] determined to examine social phenomena in terms of a system, a perspective familiar to him from his time as a biologist." (Ormerod, 2019, p. 3)

¹⁷³ "Functionalism starts from an analysis of the structures and mechanisms that hold society together, contributing to harmony and stability in a changing society" (Ormerod, 2019, p. 16)

¹⁷⁴ "Functionalism in political science focusses on the process of interaction between political institutions (e.g. parties in electoral competition or during government formation), states (e.g. during regional integration) or international organisations (e.g. when negotiating economic policies), with particular attention to growing integration of the interacting elements into a politically and thus functionally consistent system." (Berg-Schlosser et al., 2020, p. 83); In political science, functionalism emphasises the functions of political institutions, their interconnectedness and the impact on the political process. (Berg-Schlosser et al., 2020, p. 84)

¹⁷⁵ "Functionalism involves 'a gradual transfer of sovereignty from the ruler to the people ... Sovereignty cannot, in fact, be transferred effectively through a formula, only through a function. By entrusting an authority with a certain task, carrying with it command over the requisite powers and means, a slice of sovereignty is transferred from the old authority to the new' (Miller, 1971, p. 181); 'Functionalism purports to describe and to prescribe a process whereby the economic and social activities (that is, the welfare activities) of international organizations lead, eventually, to a devolution of parallel responsibilities performed by states. It traces a process of international integration, seen as a movement away from the bear-pit of competitive sovereign states towards a terminal point of world community. International organizations are the vehicles involved in this transitional process' (Miller, 1971, p. 178)

emphasis on *functionality* serving as a counterweight to national sovereignty¹⁷⁶. In neuroscience, while the general attitude of regarding the unit of study as a *system* remains, what is understood as the *solution* — that is, that towards which the operations *function* towards — is less generalisable than overall integrity or homeostasis. Oswald Steward (2000) defines the *functional neuroscience* framework as an integrative approach combining functional neuroanatomy and systems neuroscience, so as to emphasize not simply the anatomical structures of the nervous system, but their roles in enabling specific physiological and behavioral outcomes¹⁷⁷. In this framework, *functional* refers to the specific operations and contributions — the roles played — by various sub-systems (regions, anatomies, pathways...) of the nervous system towards specific behavior or processes. An intuitive way of understanding this shift towards more granular *goals* towards which the system performs *functions* is considering the “f” in fMRI, and how it relates to the exercise of a specific function (from thinking of a colour to drinking a cup of coffee)¹⁷⁸. Finally, in Philosophy of mind, *Functionalism* posits that what determines a mental state is not its internal constitution, but rather its *functional role* within the cognitive system¹⁷⁹. Thus a *function* in this context refers to the role — or causal relationships — a sub-system plays in a mental state, or at a higher level, a mental state plays in a psychological or behavioural operation. Therefore, functionalist philosophers of mind seek to understand mental states through their subsystems' roles in generating them, not their physical or material substrates. This interest in the *function* in detriment to the *substance* enables the possibility of diverse physical *realizations* of mental states (known as *Multiple Realizability*¹⁸⁰).

As per the distillation above, while *Functionalism* generally describes the effort in understanding entities systematically by what their subsystem's “do” in reference to a general or set of goals, it still relies on presuppositions derived from disciplinary inheritance. *Sociology's functionalism*, for example, starts with the analysis of existing social structures as serving necessary functions for the stability of a given society (Ormerod, 2019) — this, however, can be overly deterministic while reliant on loosely and socially defined *structures* — if one is to frame *sociological functionalism* with the sort of *multiple realizability* borrowed from philosophy of mind, one will see how it fails to account how different societies may have either similar functions fulfilled by vastly different structures, or an equilibria met via altogether different sets of enablers or restraints performing different unparalleled functions. This is an error stemming from defaulting to treating categories defined by cultural or ideological context, while not analyzing from first-principles structures based on constraint imposition, uncertainty reduction, or free energy minimisation. Each disciplinary variant of functionalism thus inherits limitations from its home domain. Biological functionalism presupposes organic substrates; sociological functionalism presupposes culturally-given structures; neuroscientific functionalism presupposes neural architectures. What is required

¹⁷⁶“In the context of Mitrany's analysis this can only mean that the doctrine of state sovereignty has interfered with the momentum of functional co-operation” (Miller, 1971, p. 182)

¹⁷⁷ “the style that has dominated the neuroanatomical teaching known as “functional neuroanatomy”, in which the focus is on regional anatomy, and how lesions at particular sites disrupt different pathways. The second is the style favored by neurophysiologists known as “systems neuroscience” in which the focus is on the physiological operation of the different functional systems. I use the term “functional neuroscience” to refer to the resulting blend.” (Steward, 2000, p. vii)

¹⁷⁸ “An fMRI can reveal what part of the brain is active during specific functions, such as lifting your arm or even just thinking about something. Researchers and physicians can use this information to better understand, diagnose, monitor, and treat various conditions.” (Pedersen, 2021, para. 5)

¹⁷⁹ “Functionalism is the doctrine that what makes something a thought, desire, pain (or any other type of mental state) depends not on its internal constitution, but solely on its function, or the role it plays, in the cognitive system of which it is a part. More precisely, functionalist theories take the identity of a mental state to be determined by its causal relations to sensory stimulations, other mental states, and behavior.” (Stanford Encyclopedia of Philosophy, 2023, para. 3)

¹⁸⁰ “In the philosophy of mind, the multiple realizability thesis contends that a single mental kind (property, state, event) can be realized by many distinct physical kinds.” (Stanford Encyclopedia of Philosophy, 2020, para. 1),

is a functionalism that presupposes nothing beyond the informational substrate itself, one where "function" is defined purely in terms of constraint propagation roles within assembly space and the path-reweighting that consolidates singularity-nodes, independent of the particular physical, biological, or social medium through which those constraints are instantiated.

The goal is to enable the cross application of Political and Cognitive Sciences terms by scrutinizing what concrete operational processes — in terms of constraint propagation, uncertainty reduction, and integrity maintenance — each term describes. Importantly, my posture is not denying the insights of classical *functionalisms*, but reframing them as situational instantiations of a more general underlying principle — one that seeks to abandon semantic structures in favor of analyzing each term by the corresponding operation, structure or architecture capacity to manipulate, restrain, process or preserve information. What is termed as "reflex" or "binding precedent" matters less than how a system channels probable states, restricts dynamics to a pre-selected set action, and feeds back constraints to sustain and execute action. Whether a cellular ion channel restricts membrane potential fluctuations or a constitutional provision restricts legislative possibilities, the informational operation—constraining state-space toward system-compatible configurations—remains functionally equivalent. This attitude is intentionally substrate-independent, enabling terminology to traverse levels without conceptual rupture, while rendering teleological "why"s into mechanistic "how"s anchored in free-energy minimization and entropy reduction. As established throughout sections 1.1–1.3 and sections 2.1–2.2, the functionalism adopted here treats governance as a configuration of roles and interactions shaping collective processing regardless of scale. The determinant question is not whether a subsystem is biological or institutional, but what role it performs in collective prediction-error minimization, higher-level Markov blanket maintenance, and the narrowing of available configurations toward states compatible with system integrity. Once a term is filtered through this lens, it seizes to be an exclusive property of any level or type of entity, recursing at multiple nested orders of organisation across substrates, albeit with different properties and instantiations. This is also important to become free from the subliminal anthropomorphic bias in analysis of systems and its associated teleological load, mainly in what is defined as "goals" or "roles" (towards them), anchoring such terms in statistical maximisation of alphabet states compatible with preferred states that minimise free energy or promote system-level maintenance.

This first-principles functionalism enables the characterisation of a concept that has threaded through the preceding sections, *Cognitive Glue*. Throughout Chapter 2, mechanisms were showcased by which agents with individual homeostatic drives become bound into higher-level cognitive systems. What unifies these mechanisms is their functional role as informational media through which multiple agents' dynamic drives become mutually constrained thereby eventuating a higher-level Markov blanket that encompasses the collective. *Cognitive Glue*, then, denotes the informational substrate — whatever its physical instantiation — that binds distributed agents into integrated cognitive systems by establishing a collective-level Markov blanket whose *Cognitive Light Cone* (2.4) exceeds those of its constituent agents. At the cellular level, bioelectric signaling through gap junctions serves this function, while past the *human as an agent* level, language and culture constitute Cognitive Glue with their own assembly indexes as historically accumulated constraint-cascades. Cognitive Glue being thus any informational medium that creates conditional dependencies between agent states, such that knowing one agent's state

reduces uncertainty about others' states, thereby producing a system¹⁸¹ with the joint distribution of agents concentrated on a strictly smaller alphabet than the Cartesian product of their individual alphabets and therefore enabling the emergence of a higher-level Markov blanket whose sensory and active states encompass the collective.

Having established *information* as a *reductionist substrate* to encompass both cognitive and governance analysis, I would like to dispel any impressions of it as an immaterial abstraction. Information is inseparably tied to the physical media that can carry, store, *interpret*¹⁸² and transmit it¹⁸³, as Karl Deutch (1966, p.84) put it, "Information has physical, "material" reality; without exception, it is carried by matter-energy processes". Albeit materially grounded, I foresee how one could question whether *Information* — cross-level/scale and relational — is compatible with functionalism and reductionism, particularly given the latter's historical search for the *smallest building block* from which to derive explanations. To these, I would answer by situating an historical tension. While Democritus' atomistic legacy focuses on fundamental *substances* (part I, preamble), Pythagoras' approach emphasises that reality is best understood through structure¹⁸⁴. Today, modern structuralists similarly prioritise relational patterns and structures over any particular underlying matter¹⁸⁵. Thus, information is not incompatible with a reductionist nor functionalist approach that distills its *substrate* from composite (that is, further reducible) *materials*. In fact, it is perfectly consistent to approach these as strategic modes of explanation, rather than a doctrine to stick to the smallest ontological particle¹⁸⁶. The reductionist substrate this thesis commits to is information as departure from maximum-entropy, on whatever probability space the level of inquiry requires, neither the floor nor the ceiling of the framework is fixed in advance. Here, the underlying *Materialism* anchors information physically (that is, only material but including the forces and energy) so as to be first-principles compatible, while allowing strategic reduction to composite structures

¹⁸¹Through positive total correlation, section 1.1.3

¹⁸² Although I have in this thesis used the framework of constraint families consolidating at singularity-nodes — as developed across 1.1 (information as the measurable departure of a joint distribution from its maximum-entropy reference, in the three registers of negentropy, mutual information, and total correlation), 1.2 (constraint accumulation as path-dependent assembly memory), and 1.3 (singularity-nodes as the configurations at which the constraint family becomes non-analytic in its parameters or the information-bottleneck Lagrangian undergoes a bifurcation), with downward causation reread in 1.3.4 as conditioning of lower-level distributions on higher-level states rather than as message-passing, to derive information across nested levels of organisation, I am purposefully leaving aside the discussion (which I would like to someday take on) regarding the symbolic vs non-symbolic instantiations of information. As per the assembly-space concepts of 1.2, given new conceptualisations of problem solving as interdimensional, multivariable problem spaces (such as 'Darwin Complete' search spaces or Assembly Theory's assembly space), it is my intuition (intuition, thus still to be developed) that by framing problem spaces as such, the signal/control level of non-symbolic information gives way towards a more semantic (non-anthropomorphically, non-teleologically defined [in the 1.3.1 sense] where 'selection' is purely the statistical reweighting of probability mass over assembly paths in favour of those whose terminal configurations persist and propagate) or symbolic level of information. In the problem space framework, search is conducted through the interpretation (even if albeit purely via matter-energy processes, after all, what alternatives are there to processes outside of these?) of patterns that reflect the interplay among local constraints, space and integrity availability, and feedback mechanisms. Within this landscape, what begins as a seemingly straightforward set of matter-energy interactions can accumulate sufficient structure and context dependency to exhibit properties that we would recognize as semantic or symbolic. The system's iterative filtering and reconfiguration processes — driven by the path-reweighting of 1.3.4 — shape how interpretive tag-states, or stable codes, might be formed and transmitted, or transformed over time. Consequently, gradually its developed hierarchical layers of reference, where the rudimentary parameters of the search become capable of 'standing for' broader contextual cues or functional objectives, even if the underlying processes remain entirely rooted in the physical interactions of matter and energy, which is an interpretable construct (symbol?)

¹⁸³ "No information without an information apparatus (...) [which determines] the very capacity of physical media for carrying and storing information" (Davies & Gregersen, 2010, p. 331) — Whether speaking of neuronal spike patterns or legislative codes, each embodies a tangible medium, electrical gradients in a membrane, ink on parchment, digital pulses in silicon, or even memetic transmissions through acoustic waves, information never floats in a vacuum, it is always carried, stored, or transmitted through matter or energy abiding by physical law

¹⁸⁴"We normally take substance to be more ontically fundamental than structure, but Pythagoras took the other perspective when he declared "All is number." (Jones, 2013, p. 37)

¹⁸⁵ These structuralists give ontic priority to structure—especially in quantum physics—and downplay the "stuff" that the structures are instantiated in" (Jones, 2013, p. 37)

¹⁸⁶ "Reductionist explanation does not have to commit to entities at the irreducible level of reality, the reduction to mechanical aggregates can serve as the models for the explanation of phenomena." (Tho, 2020)

at any scale (from subatomic interactions to institutional documentation) by focusing on its *informational potential* as elaborated on the sections above, mindful that different informational embodiments will yield different results and carry different properties and characteristics. Ion channels, gap junctions, neurotransmitters, transistors and resistors create voltage-gated current conductances (Levin, 2023) that could also be achieved by other unknown potential assemblies of whatever scale. What all these physical architectures would have in common is the *informational potential* for running *truth tables* and *logic functions* — these are a type of *informational substrate* common across scales, materials and levels that can be used as *informational compound reduction*¹⁸⁷. The usefulness of such reductionist explanation lies in its ability to model phenomena in mechanically aggregated units — such as agents, subsystems, or institutions — by distilling their *interactions* — thus yielding information — without positing that these are the ultimate building blocks of reality. It is less about individual systems (quarks, electrons or institutions) as self-contained entities, and more focused on the architecture of constraints as informational flows they enable. This *informational materialism* ought to also function as a counter to those who might ask if something *relational* is truly a (reductionist) substrate. Grounding informational first-principles in physics and statistics, yet focusing on an emergent *informational reduction* is no different than other relational physical entities such as voltage, which, by definition, measures electric potential *difference* between *two* points. The fact that it is relational by no means distances it from the inherent characteristics of its underlying physical makeup. One can “zoom in” to the molecular, transistor, or subatomic scale when necessary or useful to the informational characteristics being derived, while also allowing one to “zoom out” to the level of functional informational patterns if there’s more explanatory power there, while ideally contemplating both. Such grounding secures a common surface for these zoom lenses, ensuring that whether investigating neural circuits or policy frameworks the analysis of *informational* patterns remains tethered to the concrete processes of matter, energy, forces and probability.

Given that information is *physically real* yet functionally definable, then identical informational functions can be performed by different physical substrates. The constraint propagation that occurs through bioelectric networks in cellular systems can occur through documentary networks in governance systems functionally identically with respect to their roles in reducing uncertainty and maintaining system integrity. By contending that cognition and adaptive behavior do not depend on neural substrates, the notion that these capacities are inseparable from any particular biological solution is dispelled. This opens pathways into astrobiology, synthetic biology (as in Levin’s work on Xenobots [Blackiston et al., 2021]), and computational entities, but most critically for this thesis, it enables conceptualising orders, levels, and scales not traditionally understood as cognitive or agential to include governance assemblies.

This perspective is implicitly shared by Minsky’s (Minsky, 1986) *Ship of Theseus* like analogy, emphasising the idea that cognition arises from the organisation and interaction of parts, rather than their specific nature:

“Suppose I had once borrowed your boat and, secretly, replaced each board with a similar but different one. Then, later, when I brought it back, did I return your boat to you? What kind of question is that? It’s really not about boats at all, but about what

¹⁸⁷ The fascinating fact that *nature* gives such informational operations for free, i.e. evolution need not *invent* or *develop* logic gates, but rather seem to evolve the structures that tap or take advantage of the computational fabric of physics, inspires in me great awe as to its nature, in some esoteric way begging questions of simulation, *function*, or cognition

people mean by “same.” For “same” is never absolute but always a matter of degree. If I had merely changed one plank, we’d all agree that it’s still your boat—but after all its parts are changed, we’re not so sure of its identity. In any case, we do not doubt that the second boat will behave in much the same way—to the extent that all those substituted boards are suitably equivalent.

What has this to do with brains? Well, now suppose that we could replace each of your brain cells with a specially designed computer chip that performs the same functions, and then suppose that we interconnect these devices just as your brain cells are connected. If we put it in the same environment, this new machine would reproduce the same processes as those within your brain”

Part II

En route

Chapter 3

Humans in Assembly

3.1 The Trajectory Continues: Distributed Cognition and the Scoping Problem

To illustrate the journey leading to this thesis, and the gap in the state of the art that it attempts to cover, I'll here revisit how cognitive science has historically exteriorised cognition, and, in doing so, explain why it enables — within the framework laid out above — the search for “cognitive glues” that bind multiple human agents, while situating it in the assembly index leading up to beyond human assemblies:

The manner in which Hutchins's (1995a) widened Cognitive Science is illustrated with his reasoning that “a complete theory of individual human memory would not be sufficient to understand that which we wish to understand[,] because so much of the memory function takes place outside the individual” (p. 286).

How can it be that exhausting the study of individual memory mechanisms will not yield a complete theory of memory? The answer parallels the cellular examples explored in Chapter 2. Like no analysis of individual cellular memory mechanisms can explain tissue-level pattern memory¹⁸⁸, no analysis of individual human memory mechanisms can explain memory function that emerge from cultural-technological coordination. Consider looking strictly at the memory-output of a modern Danish citizen and of an indigenous Papua New Guinea Melanesian. If you ask them both to report on the exact events of the same day ten years ago, you will have two very distinct answers in accuracy. The Danish citizen is likely to have a social media account to log and journal events; also, is likely to have a cell phone camera at all times and a photo scroll to revisit; even a Danish luddite might keep a diary, save receipts, or at least have a banking balance statement to situate his whereabouts by proxy of his transactions and purchases. While the Melanesian indigenous, more profoundly than all of his Danish counterpart advantages, doesn't even have culturally available the concept of a calendar, from which to count ten exact years, or even know the exact current date (Telban, 2015). If this memory task aims to recall and retrieve sequential and precise events from a narrow, specific period, then our Danish has an overwhelmingly improved memory — compared to the Melanesian — as per the task's output. However this difference cannot be understood or accounted for by placing their both isolated hippocampi in a jar.

The same discrepancy is anticipated to be found between a Papua New Guinea Melanesian inscribed within a culture and his isolated *feral* twin. Cultural artefacts — far more subtle and nuanced than social media or balance statements — provide *structural* aids to cognitive tasks' outputs. Failing to acknowledge such subtle forms of cognitive offloading

¹⁸⁸ Which emerges from bioelectric network coordination, rather than residing in any single cell (section 2.2)

and structuring¹⁸⁹ leads to wrongfully endow the individual with properties which belong to a broader system¹⁹⁰. This is a manifestation of the scoping problem identified in section 1.1.3, when one fails to recognise that components of a system have positive total correlation — that knowing the state of one component reduces uncertainty about the states of others — one mistakes properties of the joint distribution for properties of any single component. This misattribution error recapitulates at the human scale the same analytical failure identified at cellular scales, attributing to individual cells morphological competencies that actually emerge from gap junction network coordination or attributing to individual neurons computational capacities that emerge from circuit-level dynamics. It is not hard to see how this misattribution becomes dangerous. Hutchins illustrates this risk by stating that *"Observed differences in level of technology between a 'technologically advanced' and a 'technologically primitive' culture will inevitably be seen as evidence of advanced and primitive minds"* (1995b, p.355).

Individual capacity is influenced by dynamics whose joint distribution registers system-level structure invisible at the component level, that is, the hallmark of systems with positive total correlation¹⁹¹. The variance-based intuition (Weiss's original formulation) fails because variance is sign-sensitive and linear-only, while total correlation registers any form of statistical dependence among components, positive or negative, linear or non-linear. A complementary illustration shared with me by my supervisor is the *bit parity* function. In a multi-bit string, the parity bit's value is fixed by the XOR of all input bits, and inspecting any single bit in isolation tells one nothing about it, that is, the parity is a property of the joint configuration that simply does not reside in any component taken alone. Bit parity captures, in compact intuitive form, the same fact that total correlation captures formally, that system-level properties cannot be read off the marginals of constituent components, no matter how exhaustively those marginals are examined. That is, individual capacity is subject to enhancers and restrainers which can only be accounted for at the *system* level. The most transparent illustration (that will soon be developed) is that of language, which no single node in a network is responsible for creating¹⁹², is stored and tailored at the network level¹⁹³, and once developed, it actively facilitates its transformation and remapping in stacks and along with it our ability to interpret and describe the world¹⁹⁴.

Recognising that individual cognitive output is irreducibly situated within higher-level systems — that what appears as individual memory, reasoning, or decision-making often reflects distributed constraint satisfaction across cultural scaffolds and institutional architectures — provides the necessary foundation for extending cognitive analysis to the governance scale. Once individual cognition is understood as emergent from participation in distributed information-processing networks, the same analytical vocabulary developed throughout Part I becomes applicable to those networks themselves, i.e., governance systems can be examined as constraint-propagating, uncertainty-reducing,

¹⁸⁹ As further developed in section 3.2.4 below

¹⁹⁰ *"drawing of the inside/outside boundary creates the impression that individual minds operate in isolation and encourages us to mistake the properties of complex sociocultural systems for the properties of individual minds."* (Hutchins, 1995b, p. 355)

¹⁹¹ As explained in section 1.1.3

¹⁹² *"Structural coupling means that language users not only invent, adopt and align their use of linguistic conventions, but also invent, adopt and align the concepts expressed by these linguistic conventions based on the outcome of their communicative interactions. Each language user stores in his private memory an inventory of conceptual building blocks, for example a set of prototypical colors, spatial relations, types of actions, classes of objects, etc. These building blocks have scores indicating how far the language user believes these concepts may lead to successful communication. When the speaker requires a concept, for example to distinguish two objects from each other, he draws on his private inventory. When no concept is available, the speaker expands this inventory with a new concept that solves the problem at hand."* (Steels, 2011 p. 351)

¹⁹³ *"In the view I develop in this article, languaging is a distributed, nonlocal organization of process on a diversity of timescales. It is one manifestation of the property of nonlocality"* (Thibault, 2011, p.218)

¹⁹⁴ As argued and cited at the section 3.2.4

integrity-maintaining cognitive architectures, revealing how both disciplines (Cognitive and Political Science) are investigating the same underlying phenomena instantiated at different scales and through different substrates.

However, historically, cognitive science has sidelined culture and environment, treating them as peripheral¹⁹⁵ for decades after its inception. This failure, to contextualise and situate cognition, lead to research that, by solely locating *processing* in the confounds of the skull, overattributed to individual minds properties that, although observed in individuals, had critical dependencies that stretched outside the border then said to be of interest to the study of cognition¹⁹⁶. Though this approach reflects an original attitude foundational to cognitive science — in which the mind is viewed as a unified and hardwired symbol processing turing machine — Hutchins' suggested a deeper rooted diagnosis. He views it as an error in scientific scoping, consequence of the aim to settle on appropriate boundaries for units of analysis in any given discipline¹⁹⁷. Any academic — concerned with framing the scope of its analysis for a cognitive phenomena — will be forgiven by having followed Plato's advice to draw a boundary where connectivity is physically low. Settling at the inner edge of the brain is then an obvious, yet misleading, compromise¹⁹⁸. The problem, in the framework of Part I, is that the choice of boundary determines the probability space (section 1.1.1), which in turn determines what can be measured. A boundary drawn too narrowly excludes constraint-cascade dependencies that materially shape the system's joint distribution, leaving the analyst unable to register the structure they are trying to explain. This thesis argues that the same scoping problem is a factor in delimiting societal cognition today.

A further element, that constrained the scope of the traditional cognitivist paradigm, was its resource to rigid, rules based structures, or pathways, in navigation and problem solving¹⁹⁹. This rigidity parallels the limitation identified in section 2.5's discussion of bureaucratic architectures, systems that rely solely on hard protocols without accommodating dynamic drives cannot adapt to novel circumstances or exploit the adaptive interpretation capacities of their constituent agents. Prior to — although in the embryonic spirit of — Basal Cognition's notion of agencies as endowed with volition, valence, and cognitive processes²⁰⁰, there was already by the mid 90s a growing body of work cementing a notion of *Soft-assembled solutions*, substituting the prior hard-wired approach with dynamically formed, flexible in reconfiguration, and adaptable solutions. This approach was informed by efforts on various fronts: Dynamical systems theory provided a toolkit of formalisations — such as attractors, stability, and phase transitions — that allowed to explain how coherent behaviour could be articulated in systems without a centralised planning algorithm (Thelen & Smith, 1994), providing a principled alternative to rigid, rules-based programming; Ecological psychology, spearheaded by Gibson's (1979) "The Ecological Approach to Visual Perception", leveraging the aforementioned concept of *Affordances*, showed that — because behavior is situated, that is, happening in response to a specific physical and social environment — action is co-constructed from contextually dependent resources, and thus not a straightforward execution of pre-scripted plans, leading to actions

¹⁹⁵ "culture, history, context, and emotion were all set aside as problems to be addressed after a good understanding of individual cognition had been achieved." (Hutchins, 1995b, p. 353)

¹⁹⁶ "If we fail to bound the system properly, then we may attribute the right properties to the wrong system or (worse) invent the wrong properties and attribute them to the wrong system." (Hutchins, 1995b, p. 356)

¹⁹⁷ "Choosing the right boundaries for a unit of analysis is a central problem in every science and the basic approach to this problem has been in place for 2,000 years. Plato advised that one should 'carve nature at its joints'." (Hutchins, 2010)

¹⁹⁸ Cybernetics had early on put emphasis on the role of environmental loops: "Early cognitive science was shaped by the tension between reductionism and holism (...) Cyberneticists like Bateson were interested in information, but they emphasized the fact that the information loops that constitute mind extend through the body into the world." (Hutchins, 2010)

¹⁹⁹ As cited and exposed during Minsky's limitations, on Chapter 5 of this thesis.

²⁰⁰ As previously detailed in section 2.1

as fluid, situated responses as opposed to fixed sequences; Motor Control research provided concrete demonstrations of soft-assembled solutions — with movement increasingly being regarded as context-specific assemblies of many forces and inputs (“*muscle synergy*”) — and thus not being the result of fixed motor commands (Bizzi et al., 1991); while Developmental Psychology began to import these concepts to, for example, substitute Piaget’s rigid stage theory in maturation with more multi dependent and factorial *spectrums of growing* (Spencer et al., 2011), which did not unfold in a *prewritten* determined way. To this day, this approach regards cognitive processes as modular to different contexts, based on the task at hand and the situation in which it is performed. For example, instead of relying on a fixed algorithm of sequences, the brain modulates parameters like muscle stiffness based on the specific requirements of a given task²⁰¹.

This dynamically enriched new attitude towards cognition equipped Hutchins to — while analysing pilot’s tasks within an airline cockpit — consider the whole of the cockpit as a system²⁰². Hutchins’ (2013) statement that “*Distributed cognition begins with the assumption that all instances of cognition can be seen as emerging from distributed processes*”, reveals that his Distributed Cognition insight, as opposed to revealing specific types of cognition, is rather a perspective to be applied to all cognitive processes. This perspective aligns with the scale-free cognition thesis developed throughout Chapter 2, cognition is not a property exclusive to certain substrates or scales but rather an emergent characteristic of any system that actively maintains integrity through constraint propagation and free energy minimisation. What Hutchins identified empirically in airline cockpits — distributed processing across human agents and technological artefacts — instantiates the same functional architecture that Levin identified in cellular collectives and that this thesis extends to governance assemblies.

This perspective argued that cognitive processes are distributed across artefacts within a socio-technical system²⁰³, thus broadening the scope of cognitive science to include interactions within systems comprising both human and non-human elements. Akin to how the *Behavioral Theory of the Firm* (Cyert & March, 1963) proved the merit of framing the business firm as a relevant unit of investigation in Economics, so did Hutchins’ insights derive the possibility for new units of analysis in Cognitive Science. Where previous market analysis had treated firms as black-boxes — entities with perfect rationality, few variables and narrow constraints — Cyert and March opened the box and found, in their words, “*a coalition of individuals, some of them organized into subcoalitions*”²⁰⁴, whose goals were sometimes mutually incompatible, whose decision making heuristics encompassed both people and artifacts, and whose interactions could not be neatly collapsed into a single, internally consistent utility function. The firm was thereby recast as a relational, dynamic output of agents and artifacts, bounded by various incentives in decision making heuristics, that coalesce into the Firms’ actions. The economic discipline thereby gained access to a level of explanation that was, in structure, what Hutchins would later open for cognitive science, that is, the analytical possibility that a system’s behaviour can only be understood

²⁰¹ “A hallmark of skilled motor performance is behavioral flexibility — i.e., experts can not only produce a movement pattern to reliably and efficiently achieve a given task outcome, but also possess the ability to change that movement pattern to fit a new context.” (Ranganathan et al., 2020)

²⁰² “An analysis of a memory task in the cockpit of a commercial airliner shows how the cognitive properties of such distributed systems can differ radically from the cognitive properties of the individuals who inhabit them” (Hutchins, 1995a, p. 265)

²⁰³ “This article presents a theoretical framework that takes a distributed, socio-technical system rather than an individual mind as its primary unit of analysis” (Hutchins, 1995a, p. 265)

²⁰⁴ “Let us view the organization as a coalition. It is a coalition of individuals, some of them organized into subcoalitions. In a business organization the coalition members include managers, workers, stockholders, suppliers, customers, lawyers, tax collectors, regulatory agencies, etc.” (Cyert & March, 1963, Chapter 3)

once its constituent agents and the conflicts and dependencies among them are made visible. And with it a theoretical framework that bridges the gap between individual cognitive properties and system-level cognitive functions (teams, meeting rooms, documents, culture...), thus — by acknowledging how systems larger than an individual can have cognitive properties in their own right²⁰⁵ — solving the aforementioned *overattribution* problem.

It thus opened the door for Cognitive Science to both analyse how information is propagated throughout systems beyond the borders of the skull, and as aggregate novel *units*²⁰⁶. For example — in his original research of an airline cockpit — the interactions among crew and apparatus within the cabin reveals patterns of cooperation, coordination and information processing through individuals and instruments²⁰⁷, which constitute a *higher than individual human level* processing system, showcasing system-level properties irreducible to individual parts²⁰⁸. This shift from individual-centric to system-level analysis directly parallels the transition established in Chapter 2, where bioelectric networks were shown to create collective cognitive capacities exceeding those of individual cells. While gap junction networks enable cellular collectives to navigate morphological problem spaces inaccessible to isolated cells (section 2.2), distributed cognitive systems enable human collectives to navigate problem spaces inaccessible to isolated individuals. The airline cockpit, in this framing, constitutes a novel Markov blanket (section 2.4), a higher-level cognitive system whose sensory and active states encompass the collective crew-apparatus assembly, whose cognitive light cone extends beyond any individual crew member's capacity to model. The functional parallel is precise: bioelectric signals coordinate cellular agents through bowtie compression-decompression cycles (section 2.5), while verbal commands, instrument readings, and procedural protocols coordinate cockpit agents through analogous compression-decompression cycles, complex situational awareness compressed into standardised callouts, then decompressed into context-appropriate actions by receiving crew members.

This newly found flexibility in determining the scope of the system's architecture in Cognitive Science, elicited a spectrum of positions on the degree to which the individual is the loci of processing — in a continuum of more to less determinant — derived from Hutchins's conceptualisation.

At one end — the "individual-plus" perspective — regards cognition as primarily an individual phenomenon, influenced by external factors. At its most centralised, it does not challenge traditional cognitivism, rather extending it by supporting the role of external factors in individual cognition²⁰⁹, with studies conducted within this framework investigating how tools and social interactions enhance individual cognitive processes. From the other far end — to the "social-only" perspective — cognition is viewed as inherently social and irreducible

²⁰⁵ "Posing questions in this way reveals how systems that are larger than an individual may have cognitive properties in their own right that cannot be reduced to the cognitive properties of individual persons" (Hutchins, 1995a, p. 267)

²⁰⁶ "This theory is concerned with how information is propagated through a system in the form of representational states of mediating structures. These structures include internal as well as external knowledge representations, (knowledge, skills, tools, etc.)." (Hutchins & Klausen, 1996, p. 1)

²⁰⁷ "The speed card booklet is a long-term memory in the cockpit system (...) The appropriate speeds for the airplane are determined by bringing the representation of the airplane gross weight into coordination with the structure of the speed card booklet" (Hutchins, 1995, p. 276). The cockpit also contains representations, such as instruments, displays, and verbal commands: "We can directly observe the many representations that are inside the cockpit system, yet outside the heads of the pilots" (Hutchins, 1995a, p. 267).

²⁰⁸ "The analysis reveals a pattern of cooperation and coordination of actions among the crew which on one level can be seen as a structure for propagating and processing information and on another level appears as a system of activity in which shared cognition emerges as a system level property." (Hutchins & Klausen, 1996, p. 1)

²⁰⁹ "Cognition is conceptualized as an individual phenomenon, but the importance of factors external to the individual is recognized" (Moore & Rocklin, 1998, p.102)

to singularly individual psychological constructs²¹⁰. This view posits that learning, thinking, and knowing are relational activities that occur situated within social and cultural contexts²¹¹. This spread illustrates and highlights once again the scoping problem in scientific research, and the relativity of *centralisation* in relation to the level in which the analysis is performed. The framework developed in this thesis resolves this apparent dichotomy. Individual cognition and distributed cognition represent different coarse-grainings (section 1.1.4) of the same underlying probability space, each level emerges by partitioning the fine-grained microstate alphabet into equivalence classes that are functionally indistinguishable at that level. Neither level is more "real" than the other, both being legitimate descriptions whose probabilistic content is fully determined by the base distribution and the chain of coarse-graining maps that connect them. Individual cognition and distributed cognition represent different scales of observation within a continuous hierarchy of nested Markov blankets, not competing ontological claims about where cognition "really" resides. Individual human agents maintain their own cognitive light cones while simultaneously participating as components within higher-level collective cognitive systems, as individual cells maintain autonomous homeostatic drives while participating in tissue-level and organism-level cognitive architectures (section 2.2).

A third perspective, heiress of Distributed Cognition and sitting somewhat in the middle of the scale laid out above, is the Extended Mind. Introduced by Clark and Chalmers (1998), it also positions cognitive processes beyond the brain, to include interactions with external resources. Yet externalism — by relying on *coupled systems*²¹², and thus leaning towards a *cognitive centre* — focuses on the concrete instances in which the mind extends beyond the brain, instead of adopting a broad attitude towards all of cognition as encompassing interactions among elements within a system²¹³. Thus externalism posits that individuals recurrently use external elements in cognitive processes and, when they do, these become part of the individual's cognitive system²¹⁴. While this approach improves Cognitive Science's perspective on the *Informational Self*, it nevertheless fails to recognize that, emergent from these coupled individuals' systems and external to them, stand other coupled systems in relation to each other, forming a higher level corpus of coupled entities that run information on a higher level, forming irreducible goals while processing information on a scale and variables scope only eligible on its own higher order assembly. This limitation parallels the analytical constraint identified in section 2.4's discussion of task environments. Externalism extends the individual's cognitive light cone through tool coupling but does not adequately theorise the emergence of collective cognitive light cones that transcend individual boundaries entirely. Externalism thus captures local constraint-coupling but stops short of the singularity-node consolidation at which the constraint family undergoes a qualitative transition and a new collective level becomes the natural locus of analysis.

²¹⁰ "Social processes are considered as cognition (...), which cannot be reduced to individual psychological constructs. Theorists who adhere to this perspective of distributed cognition argue that a view of cognition is incomplete if it does not consider social and cultural aspects." (Moore & Rocklin, 1998, p.103)

²¹¹ "Learning, thinking, and knowing are relations among people in activity in, with, and arising from the social and culturally structured world." (Lave and Wenger, 1991, p. 51)

²¹² The authors argue that "[when] the human organism is linked with an external entity in a two-way interaction, creating a coupled system that can be seen as a cognitive system in its own right." (Clark & Chalmers, 1998, p. 9)

²¹³ "extended mind picks out a kind of cognition. In the extended mind view, mind may sometimes extend beyond the brain, and sometimes it does not. Extended cognition refers only to that subset of cognitive events that involve interaction of internal and external resources. The individuation of cases of extended mind depends on empirical claims that are grounded in a container metaphor for mind. The extension of mind is manifest in links and relations that cross the usual boundary of the mind container. Second, extended mind assumes a center in the cognitive system: the organism (or the organism's brain), which is the normal mind container with respect to which cognition can be said to extend" (Hutchins, 2013)

²¹⁴ "If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is... part of the cognitive process." (Clark & Chalmers, 1998, p. 8).

It should be noted that, as a more recent development, Extended Cognition – through the model of Extended Active Inference (Constant et al., 2020) – introduced the idea of an externally realised generative model, thus treating parts of the niche as if they encode and carry information about the organism, by mathematically modeling the environment as possessing probabilistic expectations about the organism's behaviour (symmetric to how the organism has expectations about the environment), thus yielding a "shared" predictive model distributed across brain and niche. Under this formulation – for the niche's (environment) part – the *sensations* "received" are the organism's actions upon it, and the hidden states are the internal states (or goals) that led the organism to act, so, in this sense, the environment's world model would be thus updated so as for the niche to *expect* (given the organism repeatedly acting in certain ways) organisms' outcomes. This would be understood as the niche minimising environmental free energy, diminishing *surprisal* of the niche with respect to the agent's actions (the mirror image of the agent minimising *surprisal* with respect to sensory inputs from the environment)²¹⁵. As a result, the states of the environment (durable modifications, spatial arrangements, external symbols) would come to encode *beliefs* about the likely internal states of the organism, and the environment then get structured in such a way that would reflect what the organism typically wants (or expects), thereby making the organism–niche system mutually predictable and jointly self-evidencing²¹⁶. As per Constant (et al., 2020, p. 384), "*Niche and organisms can be meaningfully studied as trying to optimize their respective models of each other*", thus reframing the agent–environment boundary in a "joint self-evidencing" manner, in which organism and environment encode each-other in a "*circle of causality, in which the niche and phenotypes are trying to learn about each other to minimize their joint free energy or surprise*" (Constant et al., 2020, p. 381). I deeply empathise with this model's underlying epistemological attitude, in its unifying approach of regarding both organism and environment as structures of patterns, regularities and interdependencies, after all, the only difference between agent and environment is one of organisation at the most foundational level. However – and silverlining how informative this approach was for my understanding of Culture as a constructed cognitive niche²¹⁷ – I resist how this active inference symmetry extended to the environment might underplay the passivity of the environment's predictive process. This, to me, becomes more obvious when pointing out the model's assumption of a mostly one-to-one mapping of environmental configurations, to a single organism's type of states or actions. More often than not, environmental signals are ambiguous, layered and overlapping, encoding mixed information about multiple factors to multiple types of organisms with multiple different types of affordances, each leading to different environmental outcomes, leading the environment's "inferred expectations" to become irreducible. That is, its *internal state* would not correspond to any one single organism's prediction due to its state encoding being a circumstantial negotiation of concurrent environmental "tailors", each with their own different internal motivations. This problematises the neat symmetry in EAI's mirrored model. Complications arise while ascribing a single *generative model* to an environment serving multiple agendas, that is, the environment would "expect" incompatible things, leading to no well-defined notion of *expectation* at all, if

²¹⁵ "Under this extended form of self-evidencing, the material layout of the niche will look as if it "learns" about organismal "beliefs" causing preferred action, in the same way as organisms' learn about environmental causes generating sensations." (Constant et al... p. 383)

²¹⁶ As per the examples given in the paper (Constant et al., 2020), As many people take a shortcut across the grass toward a café, a visible trail forms. The niche has *learned* the pattern of behavior (walking trajectory) and now "advertises" it back to agents as an *affordance*. The free energy of the niche is lowered because the environment now contains a strong statistical cue about how to achieve a preferred outcome (the path reduces the surprisal of taking a given route)

²¹⁷ More on this at the "Cognitive Glue" section of this chapter

the environment lacked the mechanisms whereby to distinguish between organisms, evidentiating the passiveness of an environment's predictive processing as so defined, and its ability to actively product and test hypothesis. Pragmatically, in active inference, an agent uses its generative model actively – it performs actions to test predictions and update its beliefs — yet an unified vision of *environment*, however, lacks an independent and co-harmonised mechanism to actively probe the agent(s). Although the symmetry is interesting — and can actually hold true in tidy single agent mirror relationships (such as an individual *human culture* as a cognitive niche environment) — it might struggle to accommodate a multitypal agential scenario with competing (some overlapping, some concurrent, some antagonistic) affordance-use and environment shaping. Thus circling back to the previous paragraph's note on externalism not currently offering a route or acknowledgement of a higher order system emerging from the lower level *pairings* (with cognitive functions beyond those available at the component level alone), while highlighting the usefulness of defining — however gradient or statistical — such environment encompassing higher level system's borders.

The resolution to these limitations lies in recognising that what Extended Active Inference models as "environment" is often better conceptualised as other agents operating at the same or different scales. When multiple organisms shape the same niche, the resulting environmental configuration encodes not a single generative model but rather the outcome of multi-agent constraint negotiation²¹⁸. The "environment's expectations" that EAI posits are better understood as emergent properties of the collective constraint satisfaction process among all agents interacting with that niche. This reframing also adds credibility and another argumentative angle to the *Governance Cognition* framework developed in this thesis: governance systems emerging also through the multi-agent constraint negotiation process itself becoming structured through cognitive glues (cultural, language, institutions, coercion enforced documentation, culturally enforced tabus...) that transform passive environmental encoding into active collective inference.

As mentioned above, Externalism bridges internal functions with external resources by employing a useful principle for functional decomposition, called the *Parity Principle*. It states that if an external process performs a role that would be recognized as cognitive if it had occurred within the brain, then that process should be considered part of the cognitive system²¹⁹. This principle's underlying assumption is that cognition is determined, above all, by function²²⁰. Failing to understand this fundamental position can lead to misunderstandings in which fine-grained implementation similarities between internal and external processes are required to be defined as cognitive²²¹, and understood correctly, it encourages the assessment of cognition without bias towards biological legacy solutions, their spatial location or distribution²²². It suggests that a myriad of mechanisms of cognition can be distributed across the brain(s), embodiments and environment²²³. Tapping into the critique made above, if strictly taken, it may seem to restrict cognitive functions to those modalities

²¹⁸ i.e., the joint distribution being shaped by overlapping constraint families whose interactions produce the post-singularity configurations described in section 1.3.3

²¹⁹ "if, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process." (Clark, 2012, p. 285)

²²⁰ Aligning with the functionalist methodology of 2.7

²²¹ Heightened emphasis on the legacy biological solutions, defining the implementation over the function.

²²² "The parity claim was specifically meant to undermine any tendency to think that the shape of the (present-day, human) inner processes sets some bar on what ought to count as part of a genuinely cognitive process." (Clark, 2012, p. 285)

²²³ "The parity principle was meant to engage our rough sense of what we might intuitively judge to belong to the domain of cognition — rather than, say, that of digestion — but to do so without the pervasive distractions of skin and skull." (Clark, 2012, p. 285)

performed by human brains, disregarding operations showcasing flexible deployment of intelligence towards goal attainment through problem solving that sit outside the variables or problem-solving spaces available to human wetware. Thus excluding both potential Artificial Intelligence, and higher level Governance cognitive operations from being regarded as cognitive. In *lato sensu* though, it opens the door to accept and be on the lookout for cognitive functions and modalities outside the skull (regardless of objectively having an *inner twin pair*). The functionalist methodology²²⁴ generalises the parity principle, with cognitive status being determined by functional role in constraint propagation, uncertainty reduction, and integrity maintenance relative to system-level goals, regardless of whether that function has any analogue in individual human neurology. This enables recognition of novel cognitive modalities at the governance scale—collective memory functions, distributed decision architectures, institutional learning mechanisms—that have no individual-level counterpart to which parity could be claimed.

Another useful concept yielded by Clark in the context of exploring the implications of externalism is that of *Cognitive Scaffolding*. It starts by recognising that Human cognition leverages external structures (symbolic and social-institutional) to achieve cognitive tasks²²⁵. As Clark puts it, "*scaffolding (...) denotes a broad class of physical, cognitive and social augmentations (...) which allow us to achieve some goal which would otherwise be beyond us. Simple examples include the use of a compass and pencil to draw a perfect circle, the role of other crew members in enabling a ship's pilot to steer a course, and the infant's ability to take its first steps only while suspended in the enabling grip of its parents.*" (Clark & Toribio, 1998, p. 22). It is the principle of diffusing human reasoning across physical and social environments into which cognitive load is outsourced²²⁶. This Scaffolding is actively built, tailored or culturally sustained by *stigmergic self-modulation*, "*the process by which intelligent brains actively structure their own external (physical and social) worlds so as to make for successful actions with less individual computation*" (Clark, 1997, 191). The journey of human progress can then be seen as one of building and maintaining ever more sophisticated structures to be paired with individual cognition, evermore enhancing its capabilities²²⁷. This scaffolding construction process instantiates at the human scale the assembly pathway dynamics established sections 1.2 and 1.2.1, each successfully stabilised scaffold becomes an assembly bottleneck — a substructure of high betweenness in the directed acyclic graph of minimum-path realisations — that enables navigation of problem spaces inaccessible to unscaffolded cognition, with the scaffold itself representing accumulated constraint-cascade memory that need not be re-derived by subsequent users.

Scaffoldings exist in a spectrum of complexity. As simple as the environment alone, in the most minimal of scaffoldings, as evolution rewards organisms to use the least costly means capable of achieving their goals²²⁸. This means leveraging the structure of the environment and body to minimise internal cognitive efforts. At the simplest level, for example, Sponges use ambient water currents to aid in feeding, reducing the need for active pumping and search strategies (Day, 2024). Whereas, towards the most sophisticated, yet

²²⁴ Committed to at 2.7

²²⁵ "Where we human beings really score is in our amazing capacities to create and maintain a variety of special external structures (symbolic and social-institutional)" (Clark, 1997, p. 179).

²²⁶ "Our brains make the world smart so that we can be dumb in peace!... it is the human brain plus these chunks of external scaffolding that finally constitutes the smart, rational inference engine we call mind" (Clark, 1997, p. 180-181).

²²⁷ "We build "designer environments" in which human reason is able to far outstrip the computational ambit of the unaugmented biological brain." (Clark, 1997, p. 191).

²²⁸ "In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned" (Clark, 1997, p. 46).

ubiquitous end of the Scaffolding spectrum in humans, there's language. Though hardly opaque²²⁹, language is a major tool through which human cognition is enhanced²³⁰, with its applications being transformative in the breadth and depth of operations it opens up in cognitive systems. While Dennett sees Language as reprogramming the brain — as a cultural and linguistic transformation, beyond a fundamental change in brain structure²³¹ — Clark adds to this view by framing language as an external resource, complementing the brain's computational abilities²³². Regardless of the evolutionary impact of language in brain-structures, and the degree to which the brain has evolved to accommodate the operations performed by it — given that, although adapted to accommodate it, the brain relies on key acquisition periods in which language is acquired through socialisation²³³ (or fails to do so otherwise) — language is nevertheless an exogenous processing strategy, a cultural artefact, yet on which cognition depends for multiple operations.

The transition from passive environmental leveraging to active cultural construction marks an assembly bottleneck²³⁴ that once crossed renders accessible entirely new possibility spaces. This framework thus positioning cognitive scaffolding not as mere enhancement of pre-existing individual capacities, but as assembly pathway exploration, with the construction of historically contingent assemblies that enable navigation of problem spaces inaccessible to unscaffolded cognition.

3.2 Assembly Post Organism

3.2.1 Culture in the Assembly Continuum

Building on the Danish-Melanesian comparison from section 3.1, consider a further illustration, into a new example of situatedness across a spectrum of cultural assembly index. In a tribe with an available cultural scaffolding niche assembly with relative low index, each trade instance is a situated negotiation between concrete goods, heavily reliant on individual negotiation skills and encompassing all variables associated with each intervenient's characteristics, each assessing through their considerations past negotiations as well as future impact of the current deal, from honor to trust and reliability. Now consider

²²⁹ It amuses me to think of language as *water* in David Foster Wallace's metaphorical gag: "an older fish (...) nods at them and says 'Morning, boys. How's the water?'" And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes "What the hell is water?" (*Farnam Street, n.d.*) — we are so accustomed to relying on language to process thought, that its background role as a tool for thought itself gets often invisible by ubiquity. As we hear ourselves through inner discourse that seems intrinsically innate, it is easy to confound language with *the self*, although no single thinker has authored the words, concepts, or syntax being used to support the information processing structured through language.

²³⁰ "Public language, I shall argue, is just such a tool — it is a species of external artefact whose current adaptive value is partially constituted by its role in re-shaping the kinds of computational space that our biological brains must negotiate in order to solve certain types of problems, or to carry out certain complex projects." (Clark & Toribio, 1998).

²³¹ "The transformation of a human brain by infestations of memes is a major alteration in the competence of that organ. As we noted, the differences in a brain whose native language is Chinese rather than English would account for huge differences in the competence of that brain, instantly recognizable in behavior, and significant in many experimental contexts. (...) These functional differences, though presumably all physically embodied in patterns of microscopic changes in the brain, are as good as invisible to neuroscientists, now and probably forever." (p. 209-210) — "why do I persist in likening human consciousness to software? Because, as I hope to show, some important and otherwise extremely puzzling features of consciousness get illuminating explanations on the hypothesis that human consciousness (1) is too recent an innovation to be hard-wired into the innate machinery, (2) is largely a product of cultural evolution that gets imparted to brains in early training" (p. 219), both quotes from Dennett, 1991.

²³² "The tack I am about to pursue likewise depicts language as an external artefact designed to complement, rather than recapitulate or transfigure, the basic processing profile we share with other animals." (Clark & Toribio, 1998, p. 26)

²³³ "From the moment they are born, children engage in social interaction (...), and a child's language development is dependent on the social environment." (Lytle & Kuhl, 2017, p. 615)

²³⁴ Where multiple independent constraint sets find common resolution at a configuration that becomes reusable across many subsequent paths. Section 1.2.1

the development of an hypothetically higher index cultural-scaffolding – dependent on the assembly pathway of the concept of objective fixed characteristic, intrinsic value, and abstract simple mathematical reasoning – in which a cultural norm eventuates in which animal goods are worth twice the weight of vegetable goods in trading. This cultural-scaffolding allows for efficient trading within a broader and more vast spectrum of participants, as it uploads the assessment of various individual variables to the inherent property of the goods themselves via cultural pruning. Drastically increasing the cultural-scaffolding assembly index so as to eventuate the cultural construction of coinage, and a whole suite of new potential services as goods become objects of trade, while abstracting further still the process of negotiation to standardised operations which make it possible to negotiate across vastly greater cultural Markov blankets (section 2.4). However one still is reliant on the ability to skillfully negotiate, which diminishes further as a further cultural-scaffolding assembly index grows to encompass the concept of markets, then further much abstracted still with systems akin to Bretton Woods and FIAT, then potentially furthermore standardised and abstracted with potential crypto applications to make compute, processing or electricity as currency, each enabling in their abstraction and standardisation towards a purer functional role. Curiosity, Sara Walker (a main of the researcher on the Assembly Theory approach) during an interview hinted at this progressive assemble index into abstraction²³⁵, and even virtualisation, as a potential tendency of arrangements with very high assembly indexes. One can think of the marketplace as an higher-order Markov blanket – a singularity node²³⁶ – that encapsulates and constrains individual transaction states while enabling new forms of interaction previously impossible at the personalisable level. The same intuition has begun to surface, although under different vocabularies, in contemporary economic thought. As Petracca and Gallagher (2020) put it while arguing for treating the market as a cognitive institution, *“it is crucial to conceptualize it in terms of judgment-related, decision-making activities performed in and by the market [that is, by its agents], which on the whole allow us to see the market as an extended problem-solving mechanism”*²³⁷. The market, in this framing, is not the aggregator of transactions, but the post-singularity macrostate whose problem-solving competencies are constituted by the coordination of its constituent agents, while remaining irreducible to any of them. The cognitive light cone (section 2.4) of the marketplace encompasses variables — price equilibria, liquidity, inflation — that no individual transaction possesses, in much the same way that the cognitive light cone of an organism encompasses variables that no individual cell possesses. Such standardisations and abstractions happen when instances of processes which seek functionally identical outcomes become institutionalised in the cultural cognitive niche. While there exist functional outcomes which serve restrictive cognitive light cones – such as seeking food — there are also, functional outcomes which only become renderable at higher order Markov Blankets, composed by lower level human agents, such as lowering inflation.

²³⁵ Spoken in an informal, spoken interview setting, Sara Mentioned: *“these things that seem really abstract are just really deep in time [i.e. have very high assembly indexes]. (...) So if you think about you as a 4 billion year old object, the things that are a part of you, like your capacity use language or think abstractly or have mathematics, are just very, you know, like deep, temporal structures. (...) With our technology, we’re now building virtual realities, right? Like things we actually call virtual reality, which required, you know, four billions years of history and a whole bunch of data to basically embed them in a computer architecture.”* (Fridman, 2024, 1:25:42; 1:26:48; 1:28:31) – semi structured, on the spot, intuitions from Walker that it is possible that a property of assembly index is abstraction and even virtualisation. This aligns with the emergence of psychology as an abstract of virtual, phenomenological state, which emerges at very high assembly indexes, and it might be that higher level assemblies (from AI to Dyson spheres) could have such *virtual* properties.

²³⁶ In the sense formalised in section 1.3.3, where the constraint family of trading interactions undergoes a qualitative transition into a stable post-transition phase admitting price-equilibrating dynamics that no individual transaction possesses

²³⁷ Petracca and Gallagher’s framing was developed independently of the assembly-theoretic apparatus deployed here, yet it converges on the same structural insight, that markets are not properly explained at the level of their individual transactions, but only at the level at which their constraint family has consolidated.

This progression — from honor bond situated barter to trustless global financial markets — is not merely an historical development, but assembly pathway traversal. Each stage represents a stable constraint configuration that, once achieved, becomes available as memory for subsequent construction. The concept of “objective fixed characteristic”, such as the suggested above (e.g. animal wort 2X vegetable), is an assembly bottleneck, a constraint configuration that multiple subsequent assemblies (coinage, markets, derivatives) can reuse without re-derivation (the concept, not the specific rule). The assembly index of “market economy” vastly exceeds the assembly index of “situated barter” because it incorporates and depends upon the accumulated constraint-cascades of abstraction, standardisation, and institutionalisation that preceded it. A “market” is not a metaphor for aggregated individual transactions, it is a genuine assembly with its own integrity requirements, its own characteristic dynamics, and its own cognitive-light-cone encompassing variables (price equilibria, liquidity conditions, inflation...) that no individual market participant directly perceives. Following the resolution of the ontological critique²³⁸, “markets” are stable post-singularity phases of the constraint family that governs interactions among agents, transactions, and informational channels, with the variables that diagnose them (price equilibrium, liquidity, inflation) being properties of the ensemble whose constraint family admits price-equilibrating dynamics, not properties any individual transaction possesses. Individual traders navigate their local transaction spaces while participating as components within a market-level Markov blanket whose systemic properties emerge from but are irreducible to their individual transactions.

3.2.2 The Abstraction Bottleneck

Abstract thinking (foundational also to system’s thinking and generalisation making) seems to be *the* fundamental assembly bottleneck in the pathway from low cultural-scaffolding assembly indexes at personalisable level societies, to the higher assembly indexes in cultural-scaffoldings that eventuate higher-level *Governance Cognitive Systems*. This bottleneck parallels the dynamics of any other bottlenecks in assembly discussed in section 1.2.1, it exhibits convergent constraint satisfaction (multiple independent cultural-cognitive demands find common resolution in abstract symbolic manipulation), reusability amplification (once available, abstraction can be incorporated into countless higher-order cultural assemblies without re-derivation), and informational compression (the vast space of context-specific reasoning patterns contracts to a functionally viable subset of generalisable rules). While certain atomic configurations are prerequisites for molecular assemblies, and Chlorophyll is a prerequisite for trees, the minimal number of recursive operations required to eventuate and make culturally available abstract thinking capabilities serve as a necessary cultural-scaffolding assembly index for the standardisation and generalisations of functional roles upon which *Governance Cognitive Systems* rely. As Ivan Kroupin and Tian Chen Zeng (2024) advance in their *preprint paper* on the role of mandatory schooling and changes in social and cognitive organisation, which – although without providing a first-principles, informational nor assembly framework – converge and add on to and illustrate the importance of standardisation and abstraction, there is ample evidence from cross-cultural cognitive studies to reinforce the idea that abstract thought is highly potentiated and in fact dependent on culture-scaffolding. Take, for

²³⁸ Offered in section 1.3.4

example “A Study of Learning among the Kpelle of Liberia” by John Gay and Michael Cole (1967), in which this experiment was made:

“The task was so constructed that the objects could be sorted according to three principles (or attributes): color, number, and form. In the first problem each subject was given 8 cards (5 inches by 4 inches) on which were pictured triangles and squares, either red or green; there were 2 or 5 on a card. These 8 cards were put before the subject in a haphazard arrangement and he was asked to sort them into two groups.” (...) “The task was almost impossibly difficult for all three groups—illiterate children, schoolchildren, and adults.” (...) two modifications were made in the experiment. First of all, we tried to make sure that the subjects understood the instructions by preparing a set of sample cards on which figures were drawn in ink. The figures were large or small dots, some were filled, some open, and located in the center or near the edge of the card. The experimenter began by saying that this pack of cards could be sorted into two groups in different ways and then proceeded to form the groups in each of the three possible ways. (...) “we prepared 8 cards identical to those described earlier, but using instead pictures of a woman beating rice, with a baby on her back, and a man carrying a bucket of water on his head, followed by a dog.” (...) “the cards could be sorted according to the picture (man-woman), color, and number. As before, the subjects were requested to sort the cards in three different ways.” (...) “no significant differences between the ability to sort the triangle-square and the ability to sort the man-woman cards.” (...) The most striking aspects of these data are the relatively small proportion of subjects who managed even a second sort of the cards and the great amount of time each sort required. The average American twelve-year-old takes one look at these cards and instantly proceeds to sort them into the three possible sets.”

(Gay & Cole, 1967, p.38-39)

Or the “Recall of Classical Syllogisms” by Sylvia Scribner (Scribner, 1975) - (both examples cited by Kroupin and Zeng) showing the instance driven, casuistic and non-abstract, direct experience mode of thinking of the Kpelle farmers:

TABLE 1 Classical Syllogisms, Study I (Kpelle)	
1.	All stores in Kpelleland are in a town. Mr. Ukatu has a store in Kpelleland. Is it in a town?
2.	All houses in Kpelleland are made of iron. My friend's house is in Kpelleland. Is it made of iron?
3.	All Kpelle men are rice farmers. Mr. Smith is not a rice farmer. Is he a Kpelle man?
4.	All women who live in this town are married. Lorpu is not married. Does she live in this town?
5.	Some people we know are not in school. All the people we know are in Liberia. Are all people in Liberia in school?
6.	Some animals we know do not have Kpelle names. All animals we know are in Kpelleland. Do all animals in Kpelleland have Kpelle names?
7.	Some kwi (Western) people are wealthy. All wealthy people are powerful. Are some kwi people powerful?
8.	Some Kpelle chiefs are children. All children are single. Are some Kpelle chiefs single?

“The following [is an] excerpt from an interview with a nonliterate Kpelle farmer[:]

(...)

All Kpelle men are rice farmers. Mr. Smith (this is a Western name) is not a rice farmer. Is he a Kpelle man?

The subject replies:

S: I don't know the man in person. I have not laid eyes on the man himself.

E: Just think about the statement.

S: If I know him in person, I can answer that question, but since I do not know him in person I cannot answer that question.

E: Try and answer from your Kpelle sense.

S: If you know a person, if a question comes up about him you are able to answer. But if you do not know the person, if a question comes up

about him, it's hard for you to answer it."

(Scribner, 1975, p.155).

"The problems represented four figures of the syllogism (AA, AE, IA, OA) each one appearing in two content versions (Table 1). In one content version, the two premises express factually true propositions leading to a conclusion not only logically correct but also factually true. For example, the logically required conclusion on Problem 1, "Yes, the store is in town," is in accord with the social fact that stores in Kpelleland are located in towns." (Scribner, 1975, p.157).

(...)

"Kpelle villagers answered 53% of the problems correctly (a chance solution rate)" (Scribner, 1975, p.159).

And this isn't a single artifact of the Kpelle, as the book mentions:

"Cole (Cole, Gay, Glick, & Sharp, 1971) employed a variety of verbal logic problems in both naturalistic and experimental settings to study inferential processes among the Kpelle in West Africa (...) Explanations traditional Kpelle farmers gave Cole for their answers were strikingly similar to those given Luria by the geographically and culturally distant Siberians. Both frequently justified their answers by appealing to fact or common knowledge: "Because that's the way things are—I know it myself" (Luria, 1971, p. 270)." (Scribner, 1975, p.154).

Likewise, several replicable studies – such as with Performance on Raven's Matrices – consistently show that increasing scores from ages 8 to 18 in children attending school but no score change in their unschooled peers. These, and others examples mentioned at Kroupin & Zeng 2024, demonstrate how generalised abstract thinking — the ability to manipulate symbolic representations independent of concrete instances — is reliant on an historicity of cultural-scaffoldings, with which *coupling* enhances and operationalises further generalisations and abstractions, in a loop that externalises such abilities in the construction of cultural entities that both rely on and further potentiate this type of thinking. Being inscribed in a cultural-scaffolding at higher assembly index, with abstractions and generalisations towards functional outcomes, nudges individuals into specialisation, and, with it, the specificity of targeted functional roles. Shifting from "member of a tribe" – which encompasses the discretionary global characteristics of all members to the general needs of the individual and the tribe – to "healer" or "carpenter", involves a route in assembly index.

The Kpelle farmer's "instance-driven, casuistic and non-abstract" reasoning doesn't represent a cognitive deficit though, in fact it can be understood as the optimal cognitive processing style to a cultural niche of lower assembly index, where generalisable abstractions provide less adaptive value than situated, contextual knowledge. The American twelve-year-old's instant categorical sorting is similarly not superior intelligence as the individual markov blanket is participating in a cultural assembly that has developed and maintained the scaffolding infrastructure for abstract manipulation, an infrastructure that took millennia of cumulative cultural construction to achieve. This finding has profound implications for understanding the Danish-Melanesian memory comparison from section 3.1. The Danish citizen's superior recall derives not from enhanced individual neural capacity but

from participation in a cultural assembly with higher index, more accumulated constraint-cascade memory available for cognitive tasks. The cultural scaffolding quite literally extends the individual's cognitive apparatus by providing external structures (calendars, digital records, linguistic-temporal concepts) that individual cognition can leverage. These structures are themselves assemblies with traceable construction pathways and measurable complexity

3.2.3 Functional Role Emergence Through Assembly

In the idiosyncratic barter based trading society, each individual is expected to trade by the default and generally. In such a circumstance, there is no role of a *trader*. As for idiosyncratic and by default trading, such is the case in, for example, leader selection or harm reduction. When functional outcomes – in the realm of higher-order system events – become uploaded into cultural-scaffolding, they will tend to standardise in such a way so as to remain functional independently of the individual characteristics of who is operating the function. Below the threshold of “leader selection abstraction” in cultural-scaffolding assembly index, the leader is discretionarily selected solely based on personalisable characteristics, personal relations and other instance based variables, yet beyond the threshold – once a system creates abstractions and standardisations of leader selection – then the role of *king*, or *president* appear as functional ones, determined by the abstractions on the scaffolding of constraints sustained culturally, selecting for individuals who might have even posed little to no chance of being selected if not for the generalisable abstraction of “hereditary king” or “constitution”²³⁹. The “role of king” or “role of president” cannot be understood as properties of individuals, they are properties of the collective assembly that creates, maintains, and interprets these role-specifications. In the framework of section 1.1.4, the role specification operates at a coarser-grained level than the individual occupant. Many distinct individuals can occupy the same macrostate (“president”) because the partition that defines the role groups together all individuals indistinguishable at the role level. The role's powers, constraints, and relational structure remain invariant across individual occupants. Individual humans occupy these roles as components within a higher-level system, just as individual cells occupy functional roles (think of sieve tube cells) as components within organism-level systems. The individual brings their particular characteristics to the role, but the role itself — its powers, constraints, and relationships to other roles — exists at the collective level and persists across individual occupants.

Up to a self-enforcing point of *abstraction perpetuation*, there is a bottleneck that depends on abstract rule manipulation. Although *many* cultural assemblies are possible in configuration space, only those passing through *abstract thought enabling* cultural-scaffolding assemblies – whatever the type of entry node configuration towards abstraction – lead to assembly paths compatible with *Governance Cognitive System's* assemblies, towards which functional and generalisable roles are a dependency. Thus, the assembly paths branching out past the *abstraction* bottleneck don't necessarily eventuate any *given assembly* as we see them today, instead, these configurational routes are subject to the fit and integrity of their entire multiscale nested agents, and, given that they are free to explore a myriad of potential configurations, a cultural-scaffolding at any given index is viable

²³⁹ This transformation can be formalised as a channel that maps individual state distributions to role-constrained distributions. Where the probability of occupying a particular role becomes increasingly determined by the cultural constraint architecture rather than individual properties as the assembly index increases.

if it promotes the integrity of the societal level Markov blanket it enables. As established by the path-reweighting selection mechanic of section 1.3.4, what matters is not which combinatorially possible cultural assemblies exist, but which actually realised paths terminate in configurations whose persistence and propagation reweight the realised distribution of cultural assemblies. Paths through fragile terminations contribute negligibly to the realised cultural-assembly distribution, while paths through robust terminations channel future cultural construction.

3.2.4 Cognitive Glue: The Binding Mechanism

Having illustrated how culture exhibits assembly properties — traceable construction pathways, bottleneck dependencies, and progressive index increases — it is now possible to address the binding mechanism by which cultural assembly creates collective Markov blankets from individual human agents.

At the human level, the search for *Cognitive Glues* binding individual cognitive light cones into higher order systems follows the same functional principle of finding pathways and mediums of communication and mutual state-space pruning, towards *alphabets*²⁴⁰ compatible with the emergence of goals irreducible to a single agent, under which lower level homeostatic drives can be nested. Humans' *cognitive glue(s)* need not be mediated by a single pathway, informational type or substrate, having distributed cognition acknowledged that different types of mediums and modalities have unique physical properties that affect how information is processed and propagated²⁴¹. Thus, by highlighting the dependencies individual cognition has on a distributed architecture of dynamic interactions between *internal* and *external* modalities²⁴², cognitive science has gained an implicit framework whereby to consider structures that allow individuals to offload free-energy minimisation processes to a common module interoperating at scale, as collective channels of state-space pruning.

"Man's nervous system does not merely enable him to acquire culture; it positively demands that he do so if it is going to function at all. Rather than culture acting only to supplement, develop, and extend organically based capacities logically and genetically prior to it, it would seem to be ingredient to those capacities themselves. A cultureless human being would probably turn out to be not an intrinsically talented, though unfulfilled ape, but a wholly mindless and consequently unworkable monstrosity."

(Geertz, 1973, p. 68)

Geertz's observation acquires precise meaning within the assembly framework. The human nervous system represents an assembly bottleneck (section 1.2.1) — a stable constraint configuration that emerged through evolutionary exploration of possibility space — whose functional integrity depends on coupling with cultural scaffolding. Like mitochondria represent an assembly bottleneck whose functional integrity depends on integration within

²⁴⁰ As understood through chapter 1's framework.

²⁴¹ "Every representational medium has physical properties that determine the availability of representations through space and time and constrain the sorts of cognitive processes required to propagate representational state into or out of that medium." (Hutchins & Klausen, 1996, p. 19)

²⁴² "The pilots continually are reading and writing, reconstituting and reconstructing the meaning and the organization of both the internal and the external representations of the speeds" (Hutchins, 1995a, p. 284)"

cellular systems (section 2.2), human neural architecture represents an assembly bottleneck whose functional integrity depends on integration within cultural systems.

Human cognition's entanglement with the environment isn't unidirectional. Paired with the *cognitive offloading* processes showcased above (section 3.1) — in which an individual delegates cognitive operations into external elements, thereby diminishing internal load — humans collectively and generationally engage in a multitude of *Cognitive Uploading* processes, in which the environment itself is actively tailored and maintained across time so as to aid the transfer or pairing of cognitive operations to the environment at scale²⁴³. From a species evolutionary and biological perspective, such *uploadings* correspond to *Cognitive niche construction*, “the process whereby organisms create and maintain cause–effect models of their niche as guides for fitness influencing behavior” (Constant et al., 2020, p. 373), a phenomena not exclusive to humans²⁴⁴.

Over time — and due to the informationally diverse and cumulatively aggregated complexity of human's constructed environment — an externally shared *generative model* consolidates in the *Uploaded* pool of common cognitive resources leveraged by those maintaining the niche. This is how this thesis — inspired by Extended Active Inference (EAI)²⁴⁵ — informationally defines *Culture*, as a shared pool of collectively maintained external cognitive scaffoldings pruning surprise²⁴⁶ in individual and collective predictive and cognitive processes. The *Cultural* environment is therefore a collectively maintained memory and toolkit of problem-solving and uncertainty reduction operations, inherited by individual humans, that binds a community to a matrix of established solutions which rescue the individual from having to find individual cognitive protocols *ab ovo*. From an individual perspective, one can leverage the concept of numbers²⁴⁷ or be reminded to pick up crops by a festivity²⁴⁸ — thus offloading cognitive tasks to societally available cognitive affordances which have been collectively constructed and maintained, or *Uploaded*, externally — yet it is from a collective perspective that Culture's role as a *generative model* — and, consequently, as higher level system's *cognitive glue* binding individual human agents towards a broader *Cognitive Light Cone* — becomes apparent. This maps directly onto assembly memory as developed in section 1.2. Following the same principle that leads assembly pathways to retain successful constraint configurations enabling higher-index constructions without

²⁴³ “the notion of offloading speaking to time scales spanning individual-level dynamics unfolding over real-time and (neuro)developmental time scales and the notion of uploading speaking to individual- and group-level dynamics unfolding over developmental and intergenerational time scales. Uploading is a stronger species of offloading” (Constant et al., 2020, p. 377); “cognitive uploading endows external states of the cognitive niche with the ability to track regularities otherwise impossible to track because they are often too complex to be learned by individual organisms (...) the notion of uploading can further be viewed as the process whereby agents produce cognitive extensions that gain independence from the specific individuals having produced them” (Constant et al., 2020, p. 378)

²⁴⁴ “earthworms optimize their generative model by encoding reliable cause–effect relationships in their environment. (...) By constructing—and inheriting—semiaquatic environments like moist soils, common earthworms might have softened selection pressures on water-balance organs (Satchell, 1983; Scott Turner, 2009). Put another way, the niche became part of common earthworms' solution space to the challenge of having water removing organs in dry environments. The niche then allowed economies of “evolutionary money” to be spent on biological adaptations (e.g., selecting for water conserving organs), thereby explaining, in part, the evolutionary trajectory having led to the current phenotype.” (Constant et al., 2020, p. 385)

²⁴⁵ “We argue that the cognitive niche—understood as an externally realized cause–effects model—can be modeled as a form of externally realized “shared” generative model that is leveraged and optimized by organisms to perform action-related adaptive cognitive functions (e.g., decision-making, navigation, foraging). The optimization and leveraging of this shared generative model, through action and perception, are what we call extended active inference (henceforth, EAI).” (Constant et al., 2020, p. 377)

²⁴⁶ In the active inference sense

²⁴⁷ Field studies of the Pirahã people (Amazon) — whose language has no words for exact quantities (not even “one”) — showcase clear limits on numerical cognition when a counting system is absent. Although capable of matching quantities in immediate perceptual tasks, they struggle with memory-based or exact arithmetic tasks beyond 2–3 items (Frank et al., 2008). Similarly, the Mundurukú of Brazil (whose language has number words only up to five and, naturally, no concept of zero) can approximate large quantities, but fail at exact calculations beyond 4–5 (Pica et al., 2004). These findings suggest that without the cultural availability — without the *upload* of “numbers” into the environmental cultural niche — people rely on very crude approximate “number sensing” and cannot easily represent exact arithmetic operations, even conceptually.

²⁴⁸ “A series of rituals to boost harvest outputs can curb pest attacks” (Manan et al., 2022)

exhaustive re-exploration of combinatorial space (section 1.2), cultural scaffoldings retain successful problem-solving configurations enabling individual humans to navigate complex possibility spaces without exhaustive re-derivation. Culture is assembly memory at the human-collective scale.

Individually pairing into *Uploaded* collective cognitive functions carries two implications: “(i) *the increase in performance enabled by the outsourcing of the computation to the niche* and (ii) *the absence of an adaptation due to niche construction*” (Constant et al., 2020, p. 385). That is, tapping into a scaffolding means both being enhanced and limited by it. Growing up leveraging the yields of an environment rich in cultural affordances means interiorising a pre-structured cognitive architecture that sculpts away cognitive degrees of freedom by adhering to an established inference framework, as seen above, even sculpting away what one cognitive architecture pays attention to. The alphabet of available cognitive micro-states undergoes a selective pruning process, constrained to the sub-set of mental states compatible with the coupling with cultural affordances. This happens structurally and beyond the mere restraints imposed by “polite society”, to encompass such fundamental mental functions as “attention”. As work conducted by Constant et al. (2021) demonstrated, individual attentional styles are dynamically constructed through cultural exposure²⁴⁹, in essence implying that the cultural niche not only scaffolds individual cognitive performance by outsourcing part of the inferential load but it also constrains the attentional states by pre-structuring the very framework through which sensory information is interpreted, thus establishing a self enforcing effect in which cultural exposure simultaneously enhances and limits cognitive flexibility, yielding attentional models that are finely tuned to the specific cognitive affordances of the surrounding cultural environment.

It is worth mentioning²⁵⁰ that cultural practices which enhance immediate unpredictability are nevertheless compatible with its active inference role, given that disruptive Cultural practices ultimately lead to reducing entropy over time, aligning with the principles of Predictive Processing²⁵¹. For example, to understand the role of a cultural artefact such as “scepticism” (which leads to more rapid successions of ideas than “dogma”, and, seemingly, higher entropy) its impact must be studied over appropriate enough temporal scales to understand their impact on overall predictability (Fernandez Velasco, 2023). This temporal consideration reveals a crucial property of higher-level assemblies, their characteristic timescales differ from constituent agent timescales. Equally, as organism-level homeostasis operates over timescales that individual cells cannot directly perceive (section 2.4), cultural-level dynamics operate over timescales that individual humans cannot directly perceive nor influence. Scepticism appears entropy-increasing at individual-lifetime scales while functioning as entropy-reducing at cultural-evolutionary scales. This multi-temporal structure characterises all nested Markov blanket architectures, each level operating within its characteristic cognitive light cone bandwidth.

²⁴⁹ By employing an active inference model of visual foraging, the authors exposed simulated agents to artefacts with varying levels of decorative complexity. They showed that as the complexity of these cultural stimuli increased, agents adopted distinct visual exploration patterns—quantified through the Vertical Index—and developed unique ‘Culturally Patterned Attention Styles’ (C-PAST) characterized by changes in the entropy of motif-transition parameters. (Constant et al., 2021)

²⁵⁰ Studying culture informationally would encompass a thesis in its own right, and exhausting its theme isn’t in the scope of this current thesis.

²⁵¹ “Dynamic systems showing autovivitation (i.e. momentary increases in entropy) decrease entropy over time, and disruptive practices and learning practices are the mechanism through which these dynamics are instantiated” (Fernandez Velasco, 2023).

By pruning the available cognitive degrees of freedom across individuals²⁵² – selecting for architectures compatible with this common toolkit of scaffoldings afforded by culture – each individual interaction is eased by being contextualised within the same dependencies enforced by the cultural model²⁵³. As previously cited at footnote 140, “*the least surprising object in the world is a copy of yourself*”, so culture – thus informationally defined – is a great interpersonal diminisher of free-energy in social interactions by providing a common minimal contextual ground, enabling all sort of predictive affinities across variables between individuals that allow the formation of societal human-level Markov Blankets. The alphabet of individual state-spaces is thus diminished in a manner compatible with the emergence of higher level Markov blankets. The downward influence the higher level node exerts on lower level configurations, where higher-level constraint families project downward as conditioning that reweights the lower-level distribution, specifies the lower level human configurations that decrease overall human collective entropy. These constraints are enforced by same level dynamics which emerge from cultural artifacts — taboos, traditions, norms, legal frameworks, etiquette, status and so on²⁵⁴... — narrowing the range of possible interactions to those that are intelligible, acceptable and predictable, drastically diminishing random, erratic, or mutually incoherent actions. When two individuals share cultural scaffolding, knowing one's state reduces uncertainty about the other's state²⁵⁵. The taboos, traditions, and norms that Geertz identified do not merely constrain individual behavior but create the statistical correlations that constitute collective-level systems. Individual humans do not lose their autonomous homeostatic drives when integrated into cultural collectives, as the same way individual cells retain their metabolic integrity when integrated into multicellular organisms (section 2.2), rather, individual drives become conditionally dependent on collective states through the informational coupling that culture provides. The assembly framework developed in Part I generates the specific prediction that when lower-level agents become integrated into higher-level Markov blankets, they exhibit a characteristic signature: simultaneous constraint and enablement. At the atomic scale, this signature is unmistakable as an unbound electron in free space possesses maximal degrees of freedom and it can occupy any position, any momentum state. When bound within an atom, these degrees of freedom collapse dramatically with the electron constrained to specific orbital configurations, specific energy levels, specific spatial distributions. Yet this very constraint enables the electron to participate in molecular chemistry — covalent bonding, electron sharing... — the entire edifice of organic chemistry becomes accessible only through the constraint of atomic binding. The electron “loses” positional freedom but “gains” access to chemical possibility spaces that simply do not exist for free electrons. Section 2.2 documented how individual cells, when freed from bioelectric coordination networks (Xenobots, Anthrobots), spontaneously explore wider behavioural repertoires—they “regain” degrees of freedom. Yet

²⁵² Rather than facing every choice or situation anew, we navigate along culturally pre-sanctioned channels, default to sets of solutions using default approaches and tools; adopt a hereditary epistemology and worldview; and are literally educated to see the world through a cultural lens: to notice certain cues, ignore others, follow established scripts, experience and catalogue stereotyped emotions... That is, fall into lines of established mental state-spaces.

²⁵³ As attested by Constant et al. (2021), one can extrapolate from the effects of culture in attention so as to understand culture as furnishing each developing mind with templates not only for attention as cited, but for interpretation, action, and a whole stock of priors and heuristics that guide how one perceives and tests reality in pursue goals, and even what kinds of nested goals (beyond integrity) are pursued.

²⁵⁴ One knows, for example, that investing effort in helping a neighbor – in a society with the cultural norm of reciprocity – means the credible expectation of support and return. While, if one violates a taboo, social sanctions are expected. Such a common shared structure reduces uncertainty (surprise, predictive error) at social scale, allowing coordination and sharing at large with minimal cognitive burden on predictions. Simultaneously, low prediction error and accurate representations in a social framework of retaliation, punishment or isolation, effectively steer the individual away from state incompatible with higher order lower entropy assemblies.

²⁵⁵ Reflecting the criteria of positive total correlation, that which diagnoses systemhood, from 1.1.3

this freedom comes at the cost of access to the morphological possibility spaces that coordinated cellular collectives can navigate. The two-headed planarian, the ectopic eye, the barium-resistant regeneration—these represent problem-solving capacities that emerge only through the constraint of bioelectric integration. Individual cells "lose" behavioural autonomy but "gain" access to developmental possibility spaces that simply do not exist for isolated cells. The signature is identical: constraint at the component level enabling access to system-level possibility spaces. The substrate differs — electromagnetic forces, bioelectric signals, cultural scaffolding — but the informational architecture is invariant.

Thus the cultural niche doesn't merely provide external tools for individual cognition, but fundamentally reshape the prior distributions that govern perception-action loops in a distributed group, creating a *cultural Markov blanket* that collectively store and maintain a shared generative model. On one hand, this *cultural Markov blanket* is a supremely powerful amplifier and enabler of collective intelligence and cumulative adaptation – while providing the cognitive glue required for higher-level Societal systems – yet, on the other hand, individuals lose levels of proficiency taken from the opportunity to develop independently those skills they outsource. This is a recursive feed-forward process, given that when reliance on cultural-scaffolding increases, it necessitates a corresponding increasingly sophisticated scaffolding infrastructure to meet, distribute and execute that which is outsourced. This iterative process of cultural-scaffolding can be seen through the lens of "what are the minimal number of recursive operations required to construct *this* assembly of cultural-scaffolding?", thus representing the assembly index of cultural cognitive niche construction. As individuals delegate more cognitive operations to the cultural niche, the scaffold must evolve higher assembly indices to maintain both individual and system level integrity, which creates a positive feedback loop²⁵⁶. And so, to be able to be generally accessible and executable, the more generalisable and abstract the cultural niche needs to get, so as to permeate across individuals and situations without overfitting.

The non-fit alternative is to loose the property of being cognitive-glue²⁵⁷ thus fragmenting the societal Markov Blanket. Loose *coupling* properties so that the offloading of cognitive tasks is no longer feasible; not promoting either the integrity or reproduction of lower-level agents and thus loosing humans beyond threshold of sustaining the cultural-scaffolding; or any other fit test that compromises integrity. Following the path-reweighting mechanic of section 1.3.4, cultural assemblies whose constituent configurations fail to persist and propagate contribute negligibly to the realised distribution of cultural-assembly paths, what gets reused is what gets selected, regardless of any teleological narrative one might impose on the process. It is even viable — though contingent and precarious — to imagine a societal configuration wherein a high assembly index cultural-scaffolding maintains extraordinary resilience and systemic coherence at the societal level, while simultaneously producing significant turnover, alienation, or even chronic malaise at the level of individual human agents. The cultural scaffolding is abstract, generalised, and modular enough to absorb and replace individuals without compromising its higher-level function or integrity, similar to how sieve tube cells are sacrificed in the service of plant-wide sugar transport. The cost of such stability is borne by the lower-level agents, whose internal homeostatic drives might be persistently undermined by the cognitive pruning, role-specific constraints, and abstraction-induced estrangement required to sustain the societal-level function. Conversely, a system where the cultural niche maximises

²⁵⁶ cultural-scaffolding becomes more sophisticated, it can support more cognitive outsourcing, which in turn demands even greater sophistication.

²⁵⁷ 2.7

individual integrity, autonomy, and quality of life — perhaps by allowing more localised inference, idiosyncratic reasoning, and richer cognitive state-spaces — may result in greater variability or even fragility at the societal level, as the lack of tightly standardised, abstracted roles leads to difficulty maintaining coherent global functions across the societal Markov blanket, yet greater human-level satisfaction and well being (depending on what the higher-level goals are). This tension — between lower-level flourishing and higher-level resilience — is a political trade-off between higher-level efficiency and resilience and lower-level well being, pointing not to an optimal design but to a space of difference hierarchy of values between the assembly integrity across scales. Having this trade-off be made explicit within policy making will be evermore critical for human-well being as computational and AI driven policy-making will enable perfect higher-level stability with awful human-level impacts. The constraint-enablement signature that diagnoses assembly integration thus carries a corollary, the constraints that enable system-level possibilities may conflict with component-level flourishing. The challenge is not eliminating tension between levels (which would require eliminating nested agency) but rather designing constraint architectures that align level-specific integrity requirements to the degree possible while acknowledging irreducible trade-offs. This is a big part of *governance*'s domain, the explicit coordination of multi-level constraint architectures and their implications across levels.

With this in mind, stating that cultural-scaffoldings' assembly indexes have a bottleneck on abstract rule manipulation isn't being prescriptive about cultural-scaffolding types, nor to appraise or deprecate them based on their index. I've argued at length on this thesis about how the alternative to understanding cognitive-scaffolding is misattributing to one agent the properties of many (which, very easily, can lead to horrid and biased conclusions). Assembly index refers not to a qualitative judgement (there's no "higher = better"), but rather to the understanding of empirical minimal number of recursive operations required to construct said cultural-scaffolding. Harari, Y. N. (2015) has popularised the counterintuitive notion of hunter-gatherer tribes having higher quality of life, life expectancy and nutritional intake than subsequent agricultural societies, which are nested within cultural-scaffoldings with much higher assembly index.

3.2.5 Assembly Past the Human Object

Thus "*Monarchy*" or "*The Portuguese Constitution*", are cultural artifacts, expressing higher-order assemblies, in a lineage of cultural niche construction history and index, which are pruned by viable versus non-viable variables in a fit against functional roles in diminishing higher level free-energy and integrity, itself sustained by the roles played by culture both in pruning individual cognitive state-space to culturally available ones and the diminishing of free-energy in human prediction and interaction within a shared culturally available generative model, which abstracts individual tasks towards system-level functional roles with cultural systemic dependencies. From a first-principles, informational and non-anthropocentric perspective, governance systems represent naturally selected assemblies in combinatorial landscape, configurations whose constraint families have crossed singularity-nodes (section 1.3.3) into stable post-transition phases, with selection-as-reweighting (section 1.3.4) favouring assemblies whose terminal configurations persist under perturbation and propagate forward. Configurations that just happen to be fit enough to maintain integrity, via collective free-energy minimisation plus active construction and maintenance (due to being far from equilibrium), that have survived the pruning process

of assembly space by co-constraints at their lower-levels which maintain decreased entropic states.

When framed through the *Cognitive Light Cone* approach of section 2.4, it becomes clear that individual humans often are aligned with such temporal or causal distant goals that – akin to how a single skin cell or neuron contribute to an overall system-level sensorimotor process²⁵⁸ – can barely be conceived as existing within purely self-contained homeostatic tasks, but rather as embedded in a broader higher level goal²⁵⁹. Such goals can only be understood within the context of supra-individual protocols²⁶⁰ (from simple tribal norms to complex constitutional frameworks) where higher-level cognitive roles pursue group-scale free-energy minimization and higher level system-wide homeostasis. As a neuron's firing pattern serves system-level sensorimotor functions that the neuron itself cannot perceive or model, a citizen's tax payment serves system-level fiscal functions that the citizen cannot directly perceive or model. The neuron maintains its own metabolic integrity while participating in circuit-level computations; the citizen maintains their own economic integrity while participating in fiscal-level computations. In both cases, the lower-level agent's contribution to higher-level function is mediated through constraint architectures (neural architecture, fiscal systems) that transform individual actions into collective outcomes through information processing that occurs at the collective level.

Problem spaces in which *human agency* only becomes relevant at scale, coordinated or harmonised within a protocol²⁶¹, compose societal or governance systems' *Cognitive Light Cones* that encompass causal timescales and distances far surpassing those of any individual human. This enables framing humans as *cognitive agents*, that channel and nest personal homeostatic processes within a larger *bureaucracy* field²⁶² of higher level system's goals, even to the extent of enacting — as expected from *agents* within a broader system²⁶³ — roles in high-level assemblies' goals that co-opt and rewrite individual homeostatic drives. Micheal Levin (2019, p. 24) briefly contemplates this implication — resorting to an off-beat comparison — hinting at what he poses can potentially be a “phylogenetic *cognitive transition*” (from lower to higher level *Cognitive Light Cones*), by stating that “A cabbage cannot commit suicide [while a] A human can”. This made me speculate on the scope of drives compelling the suicidal impulse of any given individual human — across lower to higher assemblies' *Cognitive Light Cones* — as a sub-system agent. I was stricken by the thought of contemplating the likelihood of an adult grown *feral* (human) *child* committing suicide. Sustaining this thought was the hypothesis of — by being separated by goals superimposed by a higher level system — the feral child's reduced cognitive light cone would default to a “pure single system egoistic” homeostasis, separated from the higher level body of societal goals (incoming as cultural and social stressors, expectations and self-worth and identity co-constructions).

²⁵⁸ “[a] Rat learning to press a lever to get a reward (...) is a collective of cells, some of which interact with the lever (skin of the paws) and some of which get the nutrient reward (intestinal cells). However, no single cell has both experiences, and it's the causal structure of the tissue network that enables credit assignment for behaviors, so that an associative memory between an action and a reward can form that belongs to the whole animal and none of its parts alone” (Levin, 2023, p. 1866 fig. 1 A)

²⁵⁹ Think of putting a piece of paper in a box or stamping a piece of paper with ink, these are two meaningless activities without the societal scale macro-process of voting and sealing official documents.

²⁶⁰ Think of a parliament deciding on a strategy, a company learning from its history, or an institution auditing compliance adherence: these are actionable behaviours, showcasing specific cognitive functions towards system level homeostasis and free-energy minimisation.

²⁶¹ Regardless of it being explicit, binding and coercive such as a constitution; or interiorised, implicitly shared and culturally enforced like a tradition or taboo.

²⁶² In the functional sense developed in section 2.5

²⁶³ “Individual cells have this capacity” (Levin, 2019) [to kill themselves, that is, of programmed death by apoptosis] (Gordeeva et al., 2004)

As I came to learn (by briefly pursuing this tangential speculation), in primitive societies, where people expend most of their energy on immediate survival tasks (such as finding shelter, procuring food or protecting kin) and higher systems' Cognitive Light Cones are largely confined to, at most, small-group homeostatic concerns — with fewer temporal, scale and type variables — the suicide rate is severely reduced²⁶⁴ and, when occurring, largely limited to “altruistic suicides”²⁶⁵. This, at a very surface level²⁶⁶, elicits two points: one, it is possible that my thought on the likelihood of suicide by a feral child being decreased by being *unbound* to higher level goals may not completely be out of place; and second, that even at the rudimentary network level, individual human drive towards homeostasis already reflect a compromise towards higher-level goal alignment: an individual agent subsuming personal integrity under a group-system's level Cognitive Light Cone to maximize the tribe's fitness. Within the Cognitive Light Cone enabled by highly sophisticated Governance protocols, as run by modern societies — constitutions, hierarchical statutes and institutions, linguistic frameworks, cultural values — burden the *individual human* with increasingly distant or abstract societal goals, composed by a high variety of *problem spaces* irreducible to a single human's causal field, that seek societal-scale free-energy minimization and systemic integrity, extending beyond and overshadowing personal continuity. How many Ukrainian nationals, living safely abroad, volunteered to risk individual homeostatic integrity towards upholding systemic national Ukrainian maintenance?

Individual humans routinely subordinate individual-level integrity to collective-level goals in ways that can only be understood if those collective-level goals are recognised as genuine, not merely as aggregations of individual preferences but as emergent properties of collective assemblies operating at their own characteristic scales. The soldier who dies defending national territory is not maximising individual homeostasis, they are participating as a component in a national-level Markov blanket whose integrity they have internalised. Recall the altruistic suicide that Campbell (1974) identified in soldier ants—individual organisms whose jaw morphology serves colony-level functions to such a degree that they cannot feed themselves — the soldier ant's integrity is subordinated to colony-level integrity through developmental mechanisms that shape its morphology. The Ukrainian volunteer's integrity is subordinated to national-level integrity through cultural mechanisms that shape their identity. As resolved in section 1.3.4, the Downward Causation critique dissolves once one recognises that higher-level constraint families project downward as conditioning on lower-level distributions, with the soldier ant's morphology and the Ukrainian volunteer's choices being both lower-level configurations whose probabilities have been reweighted by the parameter values their respective higher-level systems supply.

²⁶⁴ “suicide was virtually unknown among the Mla Bri before more permanent settlements were established” (Long, Long, & Waters, 2013); “In the only study of suicide in gypsies, Zonda and Lester (1990) studied three districts in northern Hungary and calculated a suicide rate for gypsies of only 7.1 per 1,000,000 per year as compared to a suicide rate of non-gypsies of 22.4 (...) Some of the gypsies had assimilated and lived among other Hungarians, but even those who had assimilated felt as if they were outsiders” (Rogers & Lester, 2013)

²⁶⁵ “Altruistic suicide occurs when an individual is deeply integrated into society and chooses to sacrifice his own life to benefit other people of the society” (Shoib et al., 2022); “*The principle loci of altruistic suicide are primitive societies and the modern military*” (Stack, 2004, p. 131).

²⁶⁶ It is beyond the scope of this thesis to do a thorough multivariable analysis of a problem deeply rich in breath and depth of variables, relating to internal triggers and external conditionings leading to suicide. This passage is a mere “think aloud” meditation on the implications of Levin's (2019) comment of *suicide* as a potential “phylogenetic cognitive transition” in lower to higher goal system's transition. Although my digression into suicide rates across variables has made me curious about various avenues for future research, such as the idea that, paradoxically to the “social integration” model of suicide — which claims that strong social and communal links discourage suicide by embedding individual firm in social groups from which they derive meaning — collectivist societies with a high sense of identity (ethnstates such as Japan and South Korea) have nevertheless higher than the international average rates of suicide, regardless of having in principle high communal links, possibly hinting at the higher level pressures of system Cognitive Light Cones goals. Although an avenue I never thought I would pursue, and am not attempting to do so here, it opens a door to the question of framing suicide as suicide a friction point in the interplay between an individual's personal homeostasis and the supra-individual homeostasis of society.

Regardless, the integration of human agents into higher governance goals — beyond the scope of any individual human as it comprehends a higher level Cognitive Light Cone — is a direct implication of scale-free cognition²⁶⁷. Informationally considered, a human node exists within a field of individual to collective selves — its operations seek to maintaining personal homeostasis by receiving and integrating “*beyond any single human*” signal-constraints (from constitutional frameworks, cultural imperatives, linguistic norms...) while also ripping the benefits and responding to the incentives of performing roles within a *Markov Blanket* from which it receives information, resource allocation, and a myriad of other benefits. Furthermore — attesting to the idea of true novel higher level cognition, beyond that of mere group dynamic psychology — cognition at the Governance Cognitive system level is subject to its own types of cognitive biases²⁶⁸.

3.2.6 Language as key Assembly Bottleneck in Cultural Index

While language is nested within culture as an artifact – maintained at a higher order social level, being leveraged individually to both enhance and offload cognitive tasks – its role is determinant enough to be worth writing a dedicated segment on some of its considerations. Continuing in the line of external and distributed cognition, Clark & Toribio (1998) identified six key operations enhanced or downright *enabled* by language:

1) Memory Augmentation:

Language as an external memory aid — texts, diaries, notes, cultural narratives, histories and mnemonics alike — store data, facilitate recall and ensure information availability²⁶⁹; Notice how by delegating memory tasks to a common language, a common *uploaded* memory accessible to those who speak the language is shared and upheld²⁷⁰;

2) Environmental Simplification:

Language as a simplifier of complex environments, through labels and signs, aiding in navigation and learning²⁷¹; Logically derived, at the emergent scale, the environmental simplification labels prune the way in which speakers of a given language regard the environment collectively, diminishing misalignments in perspective;

3) Coordination and Reduction of On-line Deliberation:

Explicit plans and language-based communication, aid in coordinating actions and reducing cognitive load. Planning stabilises intentions and minimising the need for constant reassessment²⁷²;

²⁶⁷ As developed throughout Chapter 2

²⁶⁸ Which will be showcased during the third part.

²⁶⁹ “Here we simply use the artefactual world of texts, diaries, notebooks and the like as a means of systematically storing large and often complex bodies of data.” (Clark & Toribio, 1998)

²⁷⁰ Think about the word “Watergate” and how the suffix “-gate” immediately [grants memory scaffolding to](https://www.theatlantic.com/politics/archive/2014/01/how-to-name-a-scandal-what-is-a-nsbp-gate-and-what-is-a-ghazi/283104/) (https://www.theatlantic.com/politics/archive/2014/01/how-to-name-a-scandal-what-is-a-nsbp-gate-and-what-is-a-ghazi/283104/)

²⁷¹ “The provision of linguistic labels for classes of perceptually presented objects can speed category learning in artificial neural networks.” (Clark & Toribio, 1998)

²⁷² “Linguistic exchange and formulation thus plays a key role in co-ordinating activities... and in reducing the amount of daily on-line deliberation in which we engage.” (Clark & Toribio, 1998)

- 4) Taming Path-dependent Learning:
Language mitigates path dependency in learning, by allowing ideas to transfer between individuals, helping overcoming individual cognitive constraints and enabling collaboration²⁷³;
- 5) Attention and Resource Allocation:
Language aids in focusing attention and managing cognitive resources. Instructions — external or internal — guide behaviour and optimise performance²⁷⁴. The social implications of attention pruning have been stated at the section above;
- 6) Data Manipulation and Representation:
Text enables complex data manipulation and organisation. Writing and interacting with text externalises thought processes, aiding in problem-solving and idea development, allowing for degrees of complexity and idea-loop and iteration otherwise impossible²⁷⁵. Thus the collective body of text produced by language speaker/readers constitute shared problem solving heuristics.

For each of these, language functions as a dynamical structure — irreducible to any given single entity — to which processes are offloaded to support or enable an array of operations that would otherwise be impossible for isolated biological brains to perform²⁷⁶. This irreducibility is the hallmark of assembly emergence. That is, the joint distribution of language users' cognitive states exhibits positive total correlation that no individual user's marginal distribution could account for, thus making language a system in the strictest informational sense. No individual human created language; no individual human maintains it; no individual human could perform the functions language enables operating in isolation. Language represents an assembly bottleneck of extraordinary index (section 1.2.1) — the minimal number of recursive operations required to construct and maintain a functioning linguistic system vastly exceeds what any individual brain could generate *de novo*. Yet once this assembly exists as available memory in the cultural niche, individual humans can leverage it for cognitive operations that would be combinatorially impossible without it.

Each of the six functions Clark and Toribio identify instantiates the bowtie compression-decompression architecture. Language compresses high-dimensional experiential states into low-dimensional symbolic representations (words, sentences, texts) that can be stored, transmitted, and decompressed by receivers into context-appropriate interpretations. The "common uploaded memory accessible to those who speak the language" functions identically to the bioelectric pattern memory in planarian regeneration—compressed instructions that distributed agents decompress through contextual interpretation.

In addition to its role as an *individual agent enhancer*, and beyond the role it plays as a *cultural restrainer* — pruning individual's alphabet state-space towards higher-level lower entropy at the societal scale — *Language* performs a further foundational constituent role in *Governance Cognitive System's embodiment*. Given that Language is (so far, although not

²⁷³ "Public language allows human cognition to be collective... transcending the path-dependent nature of individual human cognition." (Clark & Toribio, 1998)

²⁷⁴ "We write down a note to do such and such, thus creating an externalized control loop for our own future behavior."

²⁷⁵ "In these cases, I am willing to say, a good deal of actual thinking involves loops and circuits which run outside the head and through the local environment." (Clark & Toribio, 1998)

²⁷⁶ "This power of computational transformation constitutes a neglected virtue of linguistic practice. It reveals language as the ultimate upgrade: so ubiquitous it is almost invisible; so intimate, it is not clear whether it is a kind of tool or an added dimension of the user." (Clark & Toribio, 1998, p. 38)

necessarily) the code of the protocol sustaining Governance, as the default solution to coding Governances' architectures of information processing, in the form of the written laws, bills and constitutions from which governance takes form.

This foundational role previews the argument developed in subsequent chapters: governance systems are not merely administered through language but are constituted by it, like organisms are not merely supported by bioelectric networks but are constituted by the coordination those networks enable. The transition from language as individual cognitive enhancement to language as collective governance infrastructure represents a further assembly transition²⁷⁷, from cultural Markov blankets that coordinate human agents through shared meaning, to governance Markov blankets that coordinate human agents through formalised constraint architectures.

²⁷⁷Where the constraint family that governs cultural-linguistic interactions undergoes a non-analyticity producing stable post-transition macrostate structure recognisable as governance

Chapter 4

Stable Post-Human Assemblies

Having established in Chapter 3 how cultural scaffolding enables beyond human assembly through cognitive glue mechanisms, and how abstraction represents a critical bottleneck in the assembly pathway toward governance systems, this section examines the specific informational architecture through which such assemblies achieve stable iteration. The question becomes: once cultural assembly reaches sufficient index to support functional role standardisation, what informational structures crystallise these achievements into persistent, self-maintaining configurations?

4.1 Stable recipes for iteration

As established, higher-level governance roles are abstracted from individual human idiosyncrasies, standardised as functional operations and executed at a system's level towards higher order goals. Through a structure evermore removed from individual human characteristics, Governance Systems standardise structural roles and operations in organisational and bureaucratic architectures, diminishing the reliance on individual characteristics while establishing overall consistent systems, designed to function irrespectively of who occupies or performs the role²⁷⁸.

This standardisation process — where functional roles persist across individual turnover — instantiates the constraint propagation mechanisms established in Chapter 1, and like bioelectric networks maintain tissue-level patterns despite cellular replacement (section 2.2), governance architectures maintain system-level operations despite individual human succession. The informational structure enabling this persistence, is, as elaborated on the next subsections, what this thesis terms a *Constitution* understood not in its narrow legal sense, but as the codified assembly pathway for higher-order system maintenance.

4.1.1 Constitutional Code: Referring to Classical Framings

Reiterating that I am reframing jargon from first-principles in relation to its information processing functional role, so as to bridge transdisciplinary gaps – although, from the logical articulation of the previous sections throughout this thesis, I anticipate the reader ought to already stand at a vantage point that enables them to anticipate my usage of the term *Constitution*, I am nevertheless²⁷⁹ taking a few paragraphs towards the acknowledgements below.

For classical constitutional theory, particularly the Kantian²⁸⁰ one I was exposed to during Law School, *"the significance of a constitution resides in the fact that it creates a state which is bound by rational laws, integrating citizens under norms enabling the exercise of collective human freedoms, and which does not encroach upon essential spheres of private*

²⁷⁸ "While individual leaders can shape institutions, more highly developed institutions not only survive poor individual leaders but also have a system for training and recruiting new and better ones" (Fukuyama, 2011, p. 452)

²⁷⁹ And to the despair of my supervisor, who wisely advised me to work on "Minimal Viable Deliveries" with disregard for anticipating potential counterarguments – Thank you Seth Frey for all the patience, attention and lessons shared throughout this process

²⁸⁰ (Sellers & Kirste, 2023)

human integrity" (Sellers & Kirste, 2023, p.583). Within this framework, the understanding of *constitutions* is that of a *"project for institutionalizing collective ideas of, and agreements about, the common good, framing the interests of a national community at any historical moment"* (Sellers & Kirste, 2023, p.583), with the unifying emphasis across different classical frameworks resting in the freedom granting role of constitutions²⁸¹. While, in the assembly route of Constitutions, the *freedom granting* aspect of it legitimising and motivating force might have been key in eventuating, advertising and implementing it, its key structural and informational power relates less to what it grants and more on what it establishes at the code recursive, self-referential code level.

For Hegel²⁸² and Durkheim, constitutions are seen as the structural outgrowth of modernity's functional differentiation (Sellers & Kirste, 2023), a notion that intuitively overlaps with the culture in assembly index and constitutional assembly pathway described above, down to the role of functional abstraction. *"Hegel (1969/1821) understood constitutions as legal documents that are integrally shaped by historically subliminal social processes (...) thus [he] perceived constitutions as products of transpersonal, trans-sectoral patterns of functional differentiation"*, viewing constitutions as *"evolutionary outcomes of the formative trajectories of modern social construction"*. While *"Durkheim (1950) saw the rise of constitutional law as a reflection of the transposition of society onto foundations defined by functional differentiation"* (all references above from Sellers & Kirste, 2023, p. 585).

Here, the constitution is a reflection of — and response to — the need for coordination across increasingly complex and specialised social subsystems. Its legitimacy is no longer rooted in moral law or individual autonomy, but in its capacity to stabilise expectations and generate organisational coherence across temporal and sectoral differentiation. Sociological constitutionalism, as a result, trades normativity for operability, the constitution becomes a communication-enabling constraint system, ensuring feedback, memory, and coherence across otherwise decoupled domains. This sociological framing converges with the informational framework developed throughout this thesis. The "communication-enabling constraint system" Durkheim identified can be understood as an insight that can be directly mapped onto the bowtie compression-decompression architecture²⁸³, with constitutions compressing multidimensional societal coordination requirements into codified bottlenecks that distributed agents (citizens, officials, institutions...) can interpret according to their local contexts. What the classical frameworks describe phenomenologically — stabilised expectations, organisational coherence — the informational framework explains mechanistically, constraint propagation, total-correlation increase among components (section 1.1.3) and consolidation at singularity-nodes of the multi-agent constraint family (section 1.3.3).

4.1.2 Constitution, Functional Role Crystallisation, Assembly and Integrity

²⁸¹ *"Uniting different perspectives in classical constitutional analysis, however, is the idea that a constitution acquires legitimacy as it forms a reflection of a general human interest in collective freedom."* (...) *"Underlying both classical lines of theory is the claim that the constitutional state guarantees freedoms for all society, and norms codified by a constitution underpin experiences of liberty in all social spheres"* (Sellers & Kirste, 2023, p. 583)

²⁸² I would really like to revisit Hegel's thought on spirit while framing it in perspective to the higher level entities shaped by cultural-scaffolding and Governance Cognitive Systems, and the types and styles of embodiments, behaviours, attitudes and *personalities* they assume through different *Hard Protocols* and *Dynamic Drives*

²⁸³ Explained in section 2.5

Chapter 3 established cultural scaffolding as cognitive glue binding individual human agents into collective Markov blankets, pruning individual state-space toward configurations compatible with higher-level properties. Abstraction being the bottleneck — i.e., the cultural-scaffolding *assembly index* required for generalised, context-independent reasoning — for the standardisation of functional roles. Once cultural assembly crosses this threshold, roles such as "judge," "citizen," or "legislator" crystallise as abstract positions defined by their systemic function (within the higher-level markov blanket) instead of the idiosyncratic characteristics of their occupants. Through this, *Governance* systems persist across individual turnover. Presidents die, legislators retire, judges are replaced, yet the governance system maintains its characteristic operations as crystalised by the codified *role*.

Yet role crystallisation alone does not provide constitutions its distinctiveness, given that *constraint systems* — as seen in the nature of bureaucratic architectures (2.5) — standardise functional roles. What distinguishes constitutions is the specific informational property of containing within themselves the instructions for their own perpetuation and modification. Across generations of cellular turnover within an planarian bioelectric pattern, which maintains the organisms' morphology (section 2.2), the cellular state-space has no mechanism by which to change the bioelectric pattern by which it is operationalised. That is, the bioelectric pattern itself lacks mechanisms for self-modification. The same is true for DNA, which does not codify a protocol for its own change. Constitutions, by contrast, encode not merely constraints on current behaviour but procedures for generating future constraints or altering present ones, as well as meta procedures for modifying those *modifying procedures*, along with the criteria, triggers and thresholds for determining when and if the modification is warranted and *valid*. This self-referential property, the procedural constraints on constraint-creation itself, constitutes what might be termed *recursive assembly memory*, constraint configurations that include instructions for updating themselves. A further meta aspect is the crystallisation of the constitution as *role* as opposed to version, thus *functioning* as a *constraint recipe* that includes instructions for its own replication and revision, enabling the governance assembly to maintain its coherence and integrity across the turnover of individual constitutional components and iterations.

The culmination of understanding constitutions as objects with assembly index that facilitate the emergence of beyond-human governance assemblies, points toward the ontological insight that constitutions are *constitutive* beyond merely *representational*. This distinction, though subtle, carries profound implications for understanding constitutional function within the assembly framework. By enabling the creation of said higher order assemblies and assuring their maintenance and integrity, it is, after all, a constitutive problem that constitutions are concerned with, not representation, truth or justice, but viability towards higher-level systems. As Harari (2024) posits²⁸⁴, "If DNA represented reality, we could have asked questions like 'Does zebra DNA represent reality more accurately than lion DNA?' (...) These, of course, are nonsensical questions." The analogy between constitutions and DNA is more ontological than metaphorical, both representing bottleneck assemblies in the form of stable code configurations that channel future assembly exploration while providing reusable solutions to coordination problems. Constitutions do not ask "is this proposition true?" but rather "does this imposition maintain system coherence?".

²⁸⁴ Albeit in regards to another discussion, it suits perfectly with the discussion in hand. Harari is actually analysing the concept of disinformation: "A genetic mutation that inhibits the production of adrenaline reduces the fitness of a particular zebra, ultimately causing the network of cells to disintegrate, as when the zebra is killed by a lion and its trillions of cells lose connection with one another and decompose. But this kind of network failure means disintegration, not disinformation" (...) That's true of countries, political parties, and news networks" (Harari, 2024)

Constitutions, as bottleneck assemblies, provide a material and informational basis for resilience and adaptability of higher-level structures. As DNA eventuated from assembly historicity, and proved its immense viability by providing the base solution whereby a code can function as a stable and adaptable baseline route that enables a huge array of future assemblies, so do constitutions play that role.

Like a constitution, DNA does not implement itself. It requires RNA polymerase, ribosomes, and countless other molecular machines that, although encoded within DNA itself, must already be present (produced from previous rounds of gene expression) for replication to occur, that is, neither DNA nor constitutions are self-executing. As argued and cited at section 2.2, humans are emergent from DNA being dynamically applied and interpreted — DNA not merely encoding a human in a blueprint-like way — a human is a co-constructed, enacted, interpreted and dynamically solved emergent property of lower level agencies navigating dynamically problem-spaces in selection, transcription and translation in DNA expression, in a multi-scale, multi-competent, overlapping of cognitive-lightcones. And just like a constitution doesn't contain the code for creating humans — as opposed to sustaining the creation of its organs and institutions and basis for further legislation — neither does DNA contain code for water or phosphate, which are similarly previous bottlenecks in its assembly.

From an information perspective, constitutions are assemblies that crystallise viable trajectories or further assembly routes for higher-order systems. Their defining feature is not in proclaiming freedoms (a non-scarce informational operation — e.g. instituting the rule that "everyone is free forever" — maximises entropy while providing no structural solidity), but in pruning state-space by codifying recursive, self-referential, self-iterable pathways for norm articulation, feedback, error correction, adaptation, execution, and exploration, in accordance with the dominant cognitive-scaffolding and in a manner which is capable of articulating with human level Dynamic Drives.

Drawing on Beer's *viability criterion* (1981)²⁸⁵, a constitution's success lies in whether it enables the continued coherence of the system it governs, i.e. whether it keeps the informational architecture viable as an adaptive, multi-level entity. Higher level assemblies in non-equilibria systems are eventuated due to their compatibility with – and facilitation of – lower-level Dynamic Drives. Once established, these assembly structures then require themselves operations and problem-space navigation to maintain and fulfill the lower level needs of its constituent agents so as to maintain integrity. If a constitution fails to maintain operational coherence among its subsystems, loses its alignment with lower level Dynamic Drives, or stops being enforced, it ceases to be a constitution, becoming a historical artifact. This is an extension, elaboration and integrative approach in conformity with what Deutsch (1966) already emphasised as constitutions enabling the integrity of feedback loops, receptors, effectors, and control systems, as well as the ability to maintain access to historical, internal, and external information flows. These informational capacities make constitutions as an assembly into engines of assembly-index advancement and iteration, constraining the joint distribution over the system's components such that its total correlation (section 1.1.2) sits at a level high enough to preserve systemhood (section 1.1.3) yet flexible enough to incorporate self-referential modification, further enabling the emergence and persistence of stable post-singularity phases (section 1.3.3) and novel pathways for future assemblies. At its core, it is a viability-enabling informational substrate, and while traditionally it has assumed the form of a written and cultural-scaffolding grammatical

²⁸⁵ Applied within the Cybernetic framework, to contemplate firms in his seminal book "Brain of the firm" (Beer, 1981)

language written artifact, there is no reason to believe different mediums or encodings wouldn't be as or more feasible and efficient.

Although "The Nerves of Government" is much closer associated with a classic cybernetic view of state, it nevertheless makes very important points that can be interpreted as functionally enabled operations by constitutions: *"Each of these four main functions — pattern maintenance, adaptation, goal attainment, and integration — must be served to some extent by all subsystems of the society..."* (Deutsch, 1966, p. 117); *"A society or community that is to steer itself must continue to receive a full flow of three kinds of information: first, information about the world outside; second, information from the past (...); and third, information about itself and its own parts"* (p. 129); *"All self-steering networks have three basic elements: receptors, effectors, and feedback controls. What 'selfhood,' or autonomy, such simple networks have is in their feedback controls"* (p. 128). Beyond legal positivism and natural law theory — both for this Thesis Cognitive Assembly view and the cybernetic view showcased above — constitutions are understood so as to replace these binary oppositions of classical legal arguments with a more dynamic, informational and multi-variable logic, in the sense that the core arguments boil down to the viable versus non-viability of said constitution. In assembly, the kaleidoscopic context from cultural scaffolding to biological and physical milieu, in conjunction with the constitutional code constraint architecture, either being capable of performing the role of the higher-system enabling bottleneck (toward higher level systems) or not. Consequently, on the other side of the same coin, due to the multiscale architecture of nested agencies and their lower level integrity drives, higher order viability cascades into lower order dependencies. That is, each lower order agency must retain its own viability to ensure its integrity to sustain the integrity of the higher order system. Thus, another prerequisite for viable constitutions is being capable of maintaining lower level agency and integrity, as well as their correspondent cultural-scaffolding supporting said constitution's execution. Although, it is possible to conceive of a successful constitution that—although with a high individual human agent turnover or succession rate—is able to retain overall integrity, or is able to briefly sustain integrity for long enough so as to become a consistent assembly pathway for higher index assemblies that then themselves gain irreducible characteristics capable of sustaining lower level integrity.

This viability logic shines light on a fundamental tension in constitutional design that classical theories struggle to articulate and bring forth, on the trade-off between higher-level system integrity and lower-level agent flourishing identified in section 3.2.4. A constitution optimised purely for governance-level stability may impose constraints that undermine individual human well-being while a constitution optimised purely for individual autonomy may fail to achieve the coordination necessary for collective persistence. The assembly framework reveals this not as a normative, moral, liberal vs illiberal tension, but as a structural feature of nested Markov blanket architectures, and the same that appears between cellular and organismal integrity.

The viability criterion implies that much of a constitutions' success depends on mechanisms for detecting and correcting deviations from functional configurations. From an informational perspective, DNA's double helix structure — with its specific pairing rules — represents a state of remarkably low entropy compared to an unorganized mixture of its constituents. A multivariable and multidirectional set of constraints drastically restricting the otherwise astronomical state space of possible assemblies to provide a base for stable forms of code to be established and iteratively mutated with a reliable baseline solution as

modular assembly underneath. Multidirectional referring to the interplay of both bottom-up and top-down (downward causation) influences (while multivariable is self-explanatory). Bottom-up causation, as a multiplicity of molecular forces prune DNA's available micro-states to maintain its low-entropy state, such as the electrostatic nature of hydrogen bonds acting as a specificity filter enforcing the rigid A-T and G-C geometric pairing rule (Kool, 2001), and other such mutually constraining forces that out of a universe of available micro-states formations lock each-other in lower entropy states. On the other direction, downward causation operates exactly as section 1.3.4 puts forwards, with higher-level constraint families projecting downward as conditioning that reweights the lower-level distribution $P(X)$ to $P(X|Y)$, where Y is the higher-level state. For DNA, this is exemplified by the cell's state within the cell cycle, if genomic damage is detected at the G2 checkpoint before mitosis, this systemic state of alert places the cell-cycle constraint family at a different region of its phase diagram, triggering the upregulation and recruitment of repair enzymes to the sites of damage (Harrison & Haber, 2006). The molecular-biology laws governing DNA repair still hold, but the parameter values they operate under are set by the cell-level state — the path-reweighting mechanic of section 1.3.4 acting at the molecular scale.

From this perspective, it is clear to regard the four bases (A, T, C, G) as forming a constrained alphabet (section 1.1) from which biological messages are constructed, with the joint distribution over base configurations sitting at a stable post-singularity phase that admits the assembly bottlenecks (section 1.2.1) on which everything downstream depends. Prior to the establishment of such a code, the assembly index for assembling amino acids is much more unreliable and vast in state-space. The same is true of institutions and constitutions. Although it is possible to find anthropological examples of institutions in societies — or higher order human assemblies — without a formal constitution as thought of at the next subsection, it is also possible to find amino acids in meteorites (Kvenvolden et al., 1970) and being formed by idiosyncratic settings without following the assembly route of DNA²⁸⁶. However, in both these instances, the most stable, replicable, and stable route for their assembly — which will then due to these qualities enable much more easily further en-route assemblies — happen when a codified, self-referential assembly is established. The parallel extends to error-correction mechanisms. DNA possesses proofreading enzymes and repair pathways that maintain sequence fidelity across replication cycles. Constitutions possess analogous mechanisms such as judicial reviews, amendment procedures, and interpretive precedent that maintain normative coherence across application cycles. Both represent what might be termed integrity-maintaining information processing, operations that detect deviations from viable configurations and restore system-compatible states.

The constitution is external to individual citizens (who encounter it as constraint) yet internal to Governance systems (which is partially constituted by it). Much as mRNA instructions for protein building can be external to a given ribosome—from their individual reference point as a system—but then is simultaneously an integral constituent of the individual human embodiment when taking into consideration organism level reference point, so does language—external when taken the reference of the individual system—becomes an embodied constituting binding element of the higher level Governance Cognitive System, an agency at the same level of the human agent, despite being de facto nested within the latter. This dual status follows the recursive coarse-graining structure of section 1.1.4, where the higher-level macrostate (the Governance Cognitive System and its constitutional code) is

²⁸⁶Such as by the *Miller-Urey experiment*, in the early 50s, where they simulated hypothetical conditions of Earth's early atmosphere. By passing an electric spark (representing lightning) through a mixture of gases (methane, ammonia, hydrogen, and water vapor), they synthesized a variety of organic compounds, including several amino acids (Miller, 1953).

the equivalence class produced when the lower-level joint distribution (citizens, institutions, documents) is grouped into the categories the higher level treats as indistinguishable, and the higher-level probability content follows by pushforward through the chain of coarse-grainings. In this sense, studying the properties of human grammatical language in such higher-level contexts is studying the potentialities and limitations of language as the enabling code for Governance Cognitive Systems and their architectures. Given that the information processing characteristics of language, when operating at the human-level (as an external, paired Scaffolding), will likely change when performing a higher-level role (as an embodiment of Governance). An example: let us consider second-order cognitive dynamics such as self-evaluation, self-criticism, metacognition, and other meta activities in which a system sees itself as the object to be changed, calibrated, or just observed. Studying different types of Governance codes (non-grammatical language, e.g. subsymbolic) with self-referential properties, and comparing their results to protocols enabled by grammatical language (or mixed)—in their success to observe and regulate own-states—will yield frontiers on the capacity, utility, merits and limitations of grammatical language as the constitutional code.

With the distinction between constitutional constraint and its *individual* interpretation instantiating at the governance scale the hard protocol / dynamic drive architecture already identified in section 2.6. Constitutional role definitions function as hard protocols. The specific individual occupying a role maintains their dynamic drives (personal motivations, contextual judgments, interpretive choices), but these drives operate within the constitutional constraints specifying what that role can and cannot do. The difference in interpretive bandwidth between constitutional provisions illuminates this spectrum. Some provisions (procedural requirements for legislation, term limits, jurisdictional boundaries) operate with low interpretive bandwidth, their implementation varying minimally across contexts, while other provisions (due process, equal protection, reasonable regulation) operate with high interpretive bandwidth, requiring substantial contextual and system level judgment. This variation represents constitutional design's navigation of the trade-off between stability (low bandwidth provisions that constrain unpredictably) and adaptability (high bandwidth provisions that enable context-sensitive response). Successful constitutional architectures balance both, providing sufficient hard constraint to maintain system coherence while preserving sufficient interpretive space for dynamic response to novel circumstances. The judge's role is an abstract, functional, higher weighted agent as per the Governance System's protocol, illustrating the role played by the cultural dependency on high-index of assembly necessary for abstract human roles to become embedded into the architecture of Governance Cognitive Systems. Constitutions — as both forms of cultural-scaffolding and assemblies with very high index — signify a deep and transformative bottleneck in assembly. They provide a pivotal, formalised, stable and explicit assembly on how to create and maintain a self-referential higher-order assembly which entails it as part of its own embodiment. Constitutions, as bottleneck assemblies, provide a material and informational basis for resilience and adaptability of higher-level structures. Like DNA has eventuated from assembly historicity, and proved its immense viability by providing the base solution whereby a code can function as a stable and adaptable baseline route that enables a huge array of future assemblies, so do constitutions play that role.

Constitutions represent then a specific class of bottleneck assembly, one that occurs at the intersection of cultural scaffolding and functional role standardisation. Like chlorophyll for plant assemblies, or the peptide bond for protein structures, constitutions represent

convergent constraint satisfaction points where multiple coordination requirements find common resolution. So although one could with difficulty or not at all claim that the Portuguese National Constitution came about by natural selection, one can still claim that, as an assembly, it materialised and was selected for by the the same statistical reweighting that runs everywhere else in the framework, with the same statistical reweighting of probability mass over assembly paths in favour of those whose terminal configurations persist and propagate that operates at every other scale of the framework. Furthermore, and as already hinted, constitutions solidify the abstraction process that takes place in high-assembly index cultural-scaffoldings of human roles into standardised, functional categories (e.g., "citizen," "voter," "legislator," "judge," "president"). As also seen in section 2.5's information centric approach to bureaucracy, this abstraction is a crucial form of information compression for multi-agentic architectures. By defining roles and their associated responsibilities and limitations, constitutions establish a primacy in functional higher-level roles, aims and cognitive functions which turn individuals' outputs and nested goal-seeking cognition to operate within system level operations.

4.2 From Humans to Generators: Cognition at the Governance Scale

"What major facilities of communication and memory, what machinery for self-steering, what channels of intake and foci of attention, what operational reserves available for recommitment, and what major patterns of values are found in a given political system?"

Deutsch, 1966, p. 243

The constitutional bottleneck showcased above, represents the emergence of a stable and coherent, self enforced and actively constructed integrity seeking identity — following the assembly trajectory traced throughout this thesis, in which constraint propagation compresses possibility spaces in assembly exploration, while bottleneck formations crystallise reusable solutions, and agents functionally stabilise into roles fitting higher-level properties irreducible to their constituents — culminating in the Governance Cognitive System.

A new cognitive system surfaces, operating above the human Markov blanket, persisting through constituents' individual turnover, navigating assembly and cognitive spaces previously inaccessible through its own sensory apparatus, generative model, integrity drives and actions' repertoire irreducible to any single agent. Such properties belong to the higher level markov governance system blanket as than tissue-level morphological memory is reducible to individual cellular properties, or individual human memory is irreducible to any cell. The goal of this section is not to argue by analogy by non exhaustively listing parallel in categories between individual human and Governance System cognition, but – in demonstrating functional equivalence – showcase that the same informational operations that constitute cognition at cellular and individual scales constitute cognition at governance scale. That is, differences in substrate, scale, cognitive lightcone and further idiosyncratic characteristics are expressions of a common underlying architecture²⁸⁷ and thus, what matters is not whether governance systems are *made of neurons* but whether they perform the constraint propagation, uncertainty reduction, and integrity maintenance operations that define cognitive function²⁸⁸.

²⁸⁷ expressed throughout the previous chapters of this thesis

²⁸⁸ If necessary, revisit the functionalist commitment made at 2.5

Stemming from the commitment to a functionalist methodology, I substantiate a first-principles approach to Cognition tied to integrity, viability enabling restraints, assembly index, active inference and multiscale agentic competence. From this fundamental perspective alone²⁸⁹ viable Governance System assemblies would be considered Cognitive, yet I feel compelled to supplement this argument – which takes a more foundational and assembly conception of cognition – with examples or higher level characteristics of cognition established across paradigms, and how these map into the higher-order non-reducible characteristics of Governance Systems. Note that this section isn't meant to be exhaustive as much as it is meant to complement a foundational approach with a traditional one.

Paying homage to cybernetics – from which the majority of the literature originally concerned with applying dynamical systems thinking to Governance emerged – I'll kickstart by highlighting the “*Self-Controlling*” aspect of Governance Systems. The term encompasses the innovations brought about by cybernetics that pivoted away from linear mechanical approaches, as Deutsch (1966, p. 75) puts it, “*Mechanic, organismic, and historical models were based, substantially, on experiences and operations known before 1850 (...) A major change in this situation began in the 1940's. Its basis was in the new development in communications engineering, with its extensive use of self-monitoring, self-controlling, and self-steering automatic processes*”, and with these the principles of *feedback* and *control*, painting a better understanding of systems' ability to transmit and apply information to regulate their own functions. In Governance systems context, this illustrates the ability of irreducible governance assemblies to steer themselves²⁹⁰ via higher order protocols (unavailable to single agents²⁹¹), so as to self-adjust and self-regulate.

Allow me to illustrate with one of the first policy-making events I can recall. The 2008 (to 2015) financial and Eurozone crisis saw the European Central Bank's (ECB) continuously recalibrating its approach/trajectory/processes²⁹² in relation to the distance from its goals (maintaining price stability and supporting economic recovery across the Eurozone) via a stream of input variables (inflation rates, unemployment, sovereign bond spreads and yields, and further financial data). Following a sovereign debt +crises in Greece, Ireland, and Portugal²⁹³, the ECB first abandoned its traditional monetary policy tools to implement the *Securities Markets Programme*²⁹⁴ (May 2010) only to course correct once more²⁹⁵ with the *Outright Monetary Transactions* (OMT) program²⁹⁶.

This recursive, system-level goal steering behaviour exemplifies homeostatic *Self-Control* through continuous perception-error-correction loops, demonstrating not only the system's capacity to generate forward models of desired outcomes, compare these predicted states against desired targets, and iteratively minimise the discrepancy (prediction error) to the goal, but to do so in a self-steering manner. Each policy instrument – designed, approved and implemented through protocols irreducible to any single agent – function as an interaction with the environment which contains hypothesis about optimal control and

²⁸⁹ Although elaborated throughout the thesis.

²⁹⁰ “*The first principle of control is that the controller is part of the system under control.*” (Beer, 1981, p. 25) It is governance structures that enforce governance itself.

²⁹¹ Sure a president can veto, but “a president” is not an agent in a viable functioning Constitution, it is rather an abstracted functional role emergent from the protocol itself, and so is a “veto”.

²⁹² I've written broadly regarding the inadequacy of the task-environment approach in favor of a multivariable problem space approach in in chapter 2, under the A Kaleidoscope of Selves subsection (Section 2.4)

²⁹³ Thus me paying attention

²⁹⁴ purchasing government bonds directly

²⁹⁵ following Spain's 10-year bond yields at ~7.5%, Italian spreads widening, and other far from ideal variables

²⁹⁶ shifting away from quantitative easing and introducing Conditional Automaticity, where Countries meeting EFSF/ESM criteria would automatically qualify for potentially unlimited support. I was recently 18 years old – just finished with High-School – soon after applying for Law School when I first learned the name Mario Draghi as the “[Whatever it Takes](#)” headlines spread across the news.

outcome strategies (action as prediction)²⁹⁷. Interesting also how, being a Central Bank and monitoring bond spreads, functions as problem-space type of proprioceptive feedback, enabling the ECB to perceive its own "position" within the economic problem-space.

It's also striking to realise the breath of solutions enabled by the problem-space (again stressing the limitations to task-environment made above when referring to metabolic and other spaces), when considering transitioning from conventional monetary policy to OMT as shifting from first-order controls (interest rate adjustments) to second-order controls (modifying the rules governing intervention itself), like a mitochondria shifting from standard ATP production adjustments (first-order metabolic control) to triggering mitophagy or mitochondrial biogenesis programs (Palikaras, Lionaki, & Tavernarakis, 2015) (second-order control that modifies the cellular energy infrastructure itself), changing not just energy output but the machinery and rules governing energy production within the cell.

From this *Self-Controlling* nature, a key characteristic of cognitive systems – one foundationally aligned with the first-principles assembly approach taken in this thesis – can be extrapolated in relation to what the system is interpreting, acting and adapting towards. As posited throughout this thesis, the first and most fundamental drive in any viable, high assembly index, non-equilibria system is integrity. This can be accounted by both bottom up *Dynamic Drives*²⁹⁸ supported and embodied by a shared *Cultural Scaffolding*²⁹⁹ that interprets the preservation of the Governance System as strategically aligned with lower level human integrity; and *Hard Protocols*³⁰⁰, that embed the integrity of the *Constitution*³⁰¹ at the higher order level. Two examples in which this integrity is preserved at the higher order level can be found in both the German and Portuguese constitution. The German constitution³⁰² includes an "eternity clause" prohibiting any change or removal – thus implicitly disintegration and logically derived the active maintenance – of the principles stipulated from Articles 1 and 20, which in turn guarantee democracy, republicanism, social responsibility and federalism, therefore perpetuating the *Constitutional* protocol that establishes and runs it, with a higher-level irreducible self-preservation (Polzin, 2016). Similarly established internal drives for systemic integrity can also be found at Article 288° of the Portuguese constitution (Constituição da República Portuguesa, 1976, art. 288), which, by imposing limits to constitutional revision to exclude amendments that compromise national independence and or republicanism³⁰³, incur in the same "externality clause" dynamics.

In function of this primal integrity drive – in constant friction with *metabolic* and *environmental* variables across a myriad of factors affecting the multiple problem-spaces tangible and pertinent to Governance Cognitive Systems and their embodiments – a cascade of downstream drives dynamically emerge from the systemic need to preserve

²⁹⁷ with the OMT program emerging as a meta-cognitive solution that incorporated second-order learning about the system's own learning processes, but this is another paper on its own

²⁹⁸ As developed in Section 5.2

²⁹⁹ As explained and presented at Section 3.1 (Distributed Cognition / Cognitive Scaffolding) and Section 3.2.4 (Cognitive Glue: The Binding Mechanism)

³⁰⁰ As developed in Section 5.2

³⁰¹ Understood as elaborated at Section 4.1 (Stable recipes for iteration)

³⁰² i.e. the German Basic Law (GG) (Polzin, 2016)

³⁰³ Among others

operational closure³⁰⁴, minimise entropy³⁰⁵ and maintain internal coherence³⁰⁶. These are foundational preferred states from which whole *landscapes of problem-spaces* are derived. As the system selects variables from which to triangulate its distance from those preferred states, coordinates of metrics consolidate in typologies of sub-goals.

Through these variables, the Governance Cognitive System hypothesises and makes sense of the status and changes of hidden states — like the external environment or internal embodiment states of its organs or human level agents — while being blind to the dynamics taking place outside of the scope of these selected variables.

To illustrate the statement above, consider (non-exhaustively) the sub-goals nested within a Governance Cognitive System's integrity – as conditions that need to be met so as to increase the likelihood of its viability – from legitimacy³⁰⁷ and security³⁰⁸ to economic viability³⁰⁹, coherence and information integrity³¹⁰. To operationalise these requirements, Governance Systems cluster around two information pathways: the proximity in type, and literacy competence in area, of the variables being used to assess system proximity to sub-goals; and through these proximities and relations in variable typology architecture, modular and domain-specific organs are consolidated to enable responses and actions within localised domain without necessitating overall systemic ponderation nor overhaul for domain-specific actions. From the close proximity of the variables' type that assess the system's relation to economic viability, as well as the common literacy and competence requirements in interpreting and *digesting* said variables, organs such as the Ministry of Economics or the Central Bank, for example, are demarcated within their correspondent Markov Blanket³¹¹. This modular partitioning of sub-goals — and the conflicts that arise across organs holding non-aligned aspirations — had already been intuited at the firm scale by Cyert and March, who termed the mechanism *Local rationality*. As they put it, "*an organization factors its decision problems into subproblems and assigns the subproblems to subunits in the organization*", with a corresponding "*tendency for the individual subunits to deal with a limited set of problems and a limited set of goals*" (1963, Chapter 6). Two architectural consequences follow. First, "*organizations can and do operate with much weaker rules of consistency*" than perfect rationality would require, because local goals can be concurrently pursued, and sometimes even competitively pursued, until they meet a higher-level constraint. Second, where conflicts surface across the partition, the system resolves them not by collapsing the partition into a common dimension but by "*attending to different goals at different times*"³¹² — what BTF³¹³ called *Sequential attention to goals*. The

³⁰⁴ So as for the system to be renderable and operationalisable within a logic, communication system, and execution lines renderable across it. — "*An important precondition for any adaptive system is that it must enjoy some separation and autonomy from the environment—without which it would simply dissipate, dissolve, and thereby succumb to environmental dynamics. In the absence of this separation, there would be no surprise to minimize; there must be something to be surprised and something to be surprised about. In other words, there are at least two things—system and environment—and these can be disambiguated from one another. A formal way to express a separation between a system and the rest of the environment is the statistical construct of a Markov blanket*" (Parr, Pezzulo, & Friston, 2022, p.42)

³⁰⁵ So as to crystallise predictable assemblies, pathways, organs, procedures and operational gates that allow the system not to dissipate into chaos

³⁰⁶ So as not to pursue antagonistic paths, paradoxical behaviours, self-contradicting or non-viable strategies and actions

³⁰⁷ citizen consent and procedural buy-in, so as to secure human agent level embodiment, the critical assembly bottleneck

³⁰⁸ Constituency, jurisdiction and territorial integrity, and protection from internal and external threats to *Hard Protocol* execution

³⁰⁹ Foundational resource allocation, allowing for the abstraction of human roles into governance roles, ensuring resource freedom for individual agents as parts of functional abstracted protocols' roles can perform their function while still being grated resources that align with their *Dynamic Drives* (as defined in "Luhmann's Error" section)

³¹⁰ Legal logic continuity so as to allow bureaucratic reproduction, execution and trustworthiness, data interpretability and perception and execution channels

³¹¹ As defined and elaborated at the second chapter

³¹² "*The business firm is likely to resolve conflicting pressures to 'smooth production' and 'satisfy customers' by first doing one and then the other*" (Cyert & March, 1963, Chapter 6).

³¹³ *Behavioural Theory of the Firm* (Cyert & March, 1963)

example the authors give is firm-level (smoothing production versus satisfying customers), but the mechanic is general, and reappears at every nested level of organisation. Autoimmunity in immunology, in which immune cells acting on their local drive to eliminate perceived threats initiate responses detrimental to organismal health when self-antigens are mistakenly targeted (Devarajan & Chen, 2013), instantiates the same principle from the cellular side. The Eurozone trajectory traced above instantiates it from the governance side, in that the ECB's sequential prioritisation of inflation, sovereign bond spreads, and growth across the 2008–2015 cycle is *Sequential attention to goals* operating at supra-national scale. These structural signature of bounded cognitive light cones (section 2.4) coordinating across a partition whose total correlation (section 1.1.3) is positive but whose joint distribution does not admit collapse onto a single common dimension.

The relationship between variable selection and sub-goal formation reveals how governance systems construct their generative models. In section 2.1 it was detailed how generative models encode predictions about environmental states. At governance scale, these predictions crystallise into *governance rendered metrics* against which states are evaluated. A Cognitive System demonstrates *awareness* through its response to stimuli³¹⁴, as an integration of variables that affects the system's operations, provoking a change in the system's internal mechanics³¹⁵ so as to inform behaviour, in a manner that makes sense in relation to how the stimulus updates the system's world model³¹⁶. This determination is based on reliable repetition and appropriate integration over dramatic one-time incidents³¹⁷. It should also be abstracted from self-consciousness, which can be seen as an architecture of meta-stimuli response — i.e. as *awareness of awareness* and directness of attention to internally generated variables³¹⁸ — rather, awareness should be understood in relation to problem-space navigation, as the system adapts to changes caused by variable rendering (that is, putting again emphasis on the *controller* being the system itself).

For example, being economic viability a core sub-goal of systemic integrity and inflation a key economic factor – too high inflation erodes undermines social trust and threatens *Constitutional* integrity, as seen in the Weimar Republic (Galofré-Vilà, 2023), Zimbabwe between 2007 and 2008 (Kramarenko et al., 2010) or Venezuela (Pittaluga et al., 2021) – viable Governance Cognitive Systems have come to integrate in its *array of sensors*, i.e. in the variables they seek to integrate so as to adapt its generative model and inform problem-space navigation, variables such as the *Harmonised Index of Consumer Prices* (HICP).

This variable sensitivity showcases the bureaucratic compression mechanisms laid out at section 2.5. The HICP compresses multidimensional economic reality — millions of individual transactions, regional variations, sectoral differences — into a single bottleneck, necessarily discarding information³¹⁹ while enabling coordination across agents who cannot possibly model the full economic state space, while enabling system level *sensitivity*, is in relation to which self-steering towards preferable states intercepts with valence³²⁰, as the

³¹⁴ "If the system responds to this stimulus, then we say that it is aware... All that we know from our experiment is that the system is aware of a change wrought in itself" (Beer, p. 26).

³¹⁵ "Clearly what is to count as a stimulus is an interference which affects the system's operation in some way, being neither so mild as to be insignificant to the operation nor so drastic as to destroy the system itself." (Beer, p. 26)

³¹⁶ More on this on the next paragraph

³¹⁷ "Let us define awareness behaviourally, that is, by specifying the behaviour which is to count as typical of an aware system." (Beer, p. 25) "We might well train the cat to leave the room on seeing the board. If it always did so, we would abandon the idea that this was a coincidence and talk about a response to a stimulus, and an aware cat." (Beer, p. 26);

³¹⁸ Such reflection on metacognition has already been made on Section 2.5 (Cognition, Governance or principles of operating integrity?), (Sandved-Smith et al., 2021)

³¹⁹ section 2.5's discussion of compression faults

³²⁰ As elaborated at the Chapter 2 sub-section "A many self's' Self: cognition and valence all the way down"

preference or avoidance of particular stimuli or conditions, which allows non-equilibria entities to maintain integrity by navigating problem-spaces so as to steer towards preferable states. To continue the example, a $\pm 0.5\%$ buffer around 2% annual inflation is the consensual target in western central banks³²¹, this is the preferred valence state made rendable by variable sensitivity, from which the system is made aware of positive (remain) or negative (adapt) drives.

Suppose HICP³²² data show inflation at 3.1% , 2.9% , then 3.0% over three consecutive months, then the system would have a stable and predictable stream of valence from which to be made aware of distance from a preferred state, that would compel action within this problem-space. Whereas in variables lacking valence or in the absence of variables to make the system aware altogether, the system would be unable to form sensitivities that sustained both an up to date and rich world model and the drive to remain situated within preferable problem-space coordinates compatible with integrity. As a last comment, given this notion could be easily confused with valence, I would like to stress the idea that Governance Cognitive Systems integrating negative or positive valence does not directly overlap with ethical or justice values (*good* or *bad* $\neq$ positive or negative valence), rather and merely a classification of stimuli as harmful or beneficial as it translates to approximating or distancing from any said goal, regardless of its moral taint, that is, good or bad in valence relates to the distance or proximity to the desired coordinates in problem-space to a given variable in function of a goal.

Through modularity, organizations achieve a balance between stability and adaptability, essential for functioning in complex environments and aligning specialized tasks with broader, multi-functional goals (Jones, 2001, Chapter 6). These are the same modular architecture principles, eventuated by the same foundational principles, found throughout cognitive systems. Understood as the presence of semi-independent, functionally specialised substructures within a broader Cognitive System (Fodor, 1983), modules are a foundational characteristic from (traditionally said) *artificial* (Small, 1998) to (said)³²³ *natural* systems, enabling the parallelised and domain-specific processing of inputs distinct in variables and metrics — for example visual processing (Zeki & Bartels, 1998) or memory encoding and retrieval (Moraschi et al., 2020) — optimising informational throughput and reducing cross-domain interference (Barsalou, 2008). To quickly elaborate on the above's mentioning of the system's *blindness* to stimuli outside the spectrum of what can be picked up by the variables it is sensitive³²⁴ to, allow me to introduce the concept of *externalities*. If the Governance Cognitive System doesn't contain any sensibility to the variable³²⁵ "*biodiversity*", then it will have no way of accessing the *hidden* external environmental status that corresponds to this fact, remaining as blind to it as human systems to non-visible light as the lack the sensibility to process those wavelengths as variables. Consider a rudimentary Governance System with limited sensibility — i.e. with a limited scope of variables it can take into consideration or with deficient pathways to gather or deliver those variables with stability — and/or poor modeling capabilities, that is, the capacity to represent

³²¹ Most advanced-economy central banks aim for 2% annual inflation. This midpoint: Keeps inflation low enough to maintain purchasing-power stability; Keeps inflation high enough to preserve monetary policy room; Anchors long-term inflation expectations around a stable, credible norm. (Richmond Federal Reserve Bank, 2024)

³²² "The harmonised index of consumer prices (HICP) serves 2 main purposes: it is a measure of price stability for the European Central Bank's (ECB) monetary policy strategy. The Governing Council of the ECB has defined price stability in terms of the HICP for the euro area; it is used as an indicator for assessing price convergence regarding a possible accession of a country to the monetary union (euro currency)." (Eurostat, n.d.)

³²³ I think from the nature of this thesis the reader is in a good vantage point to understand my problem with the anthropocentric and arbitrary separation in the definitions between artificial and natural when it comes to intelligence and cognition

³²⁴ Further elaboration on sensitivity on this section below

³²⁵ More on what it means to represent or not a given variable coming up in this subsection

and the literacy to interpret and render the variables into coherent hypotheses about hidden states. Such a system would have a narrower horizon of elements to consider or hypotheses about the world state to test and validate, and thus, from this diminished *Cognitive Light Cone*, a smaller problem-space to act upon. However, the ontological environment a system interacts with and is inscribed at remains the same regardless of its *Cognitive Light Cone*. A Governance system with no sensibility to monetary variables is susceptible to the same viability and integrity threats posed by hyper-inflation, regardless of being unable to render their underlying dynamics. This exerts a selective pressure towards Governance systems that represent variables key for integrity and viability, a pressure that operates whether or not the system itself can perceive it, operating exactly as the path-reweighting selection mechanic of section 1.3.4, where the realised distribution over governance assemblies departs from the prior distribution in proportion to the persistence of each path's terminal configuration.

Furthermore, if there isn't a functional thalamus like architecture or solution to bind stimuli across modularities sensitive to different variable typologies, then the system is at great risk of incurring in pathologies similar in bug type to visual agnosias or further disconnection syndromes, in which although a given module is capable of processing and rendering variables partaking its *expertise*, problems in relaying or further modular interpretability make it so that at the system level said stimuli is inoperable. If a ministry of environmental affairs has neither connections nor weighting in the overall process of execution and behaviour, then although capable of rendering said variables the overall system would incur in a kind of agnosia.

Due to the scale we occupy – and the tradition of separating humans and human structures from the unfolding of events in the natural world, with a lingering implicit “human versus natural” cognitive bias in science and culture – attributing to Assembly history and index, information gain or Active Inference the natural nesting of lower level humans into higher level assemblies (which exist as higher order ontological cognitive systems with irreducible goals and properties) can be counter the enlightenment view of Man as the measurement of all things. “How are countries or institutions not artificial if we created them?” is a reaction I anticipated throughout the writing of this thesis. However, if we take the scale of a mitochondria, and judge artifice by the criteria of the question framed above, how are we to judge chlorophyll, the phospholipid bilayer or serotonin? All these are cell's *artifices*, structures whose bottleneck into existence pass through being produced by cells. This objection dissolves once the assembly framework of Chapter 1 is properly understood. Section 1.2 drilled down on how all high-index assemblies are “artificial” in the sense that they represent historically contingent constraint configurations that emerged through selection processes. In fact, section 1.3.4 made explicit that the realised assembly graph is heavily biased toward paths through robust intermediates, with bottlenecks (section 1.2.1), singularity-nodes (section 1.3.3), and selection (section 1.3.4) being three vantages of the same underlying phenomena. The question is not whether governance systems are “natural” but whether they exhibit the informational architecture characteristic of cognitive systems.

While abstracting away from individual level characteristics and idiosyncrasies, the Governance System acquires – through both by virtue of the higher-order architecture adopted and the nature of the cultural-scaffolding that interprets and enacts its *Hard Protocols* – higher-order traits irreducible to any individual human. Not only the difference in priorities, goals or variables the system is sensible to will determine its temperament, traits and attributes, but so will the architecture of functions and operations in relationship to these.

Architectural differences in formation processing yield different higher-level systems. For example, *“Dictatorial information networks are highly centralized”*³²⁶, with information *“flow[ing] to the central hub, where the most important decisions are made”*. This with architectural structure come various downstream implications, for example, as Harari (2024) notes, *“In theory, a highly centralized information network could try to maintain strong self-correcting mechanisms, like independent courts and elected legislative bodies. But if they functioned well, these would challenge the central authority and thereby decentralize the information network”*, thereby offsetting the foundational architectural design of the system’s highly centralised protocol.

Likewise, such highly non-distributed architectures *“want[s] all information to pass through the central hub and doesn’t want any independent institutions making decisions on their own”*³²⁷, given that this would likewise compromise the integrity of the centralised architecture. Consequently, *“Since they don’t allow information to flow anywhere except through the official channels, if the official channels are blocked, the information cannot find an alternative means of transmission”*, leading to an overall system and a corresponding embodiment with a highly skewed and nonoptimal prediction error integration system, leading to an underdeveloped generative model. *“Democracy, in contrast, is a distributed information network”*, with *“ongoing conversation[s] between diverse information nodes”* as *“Information freely circulates between private businesses, private media organizations, municipalities, sports associations, charities, families, and individuals without ever passing through the office of a government minister”*. This is an extreme and narrow example of how different Governance Cognitive System’s architectures of information processing will lead to different behaviours and characteristics – representing a principle that holds across all morphology changes in Governance Systems, not unlike how small neurotransmitter changes or neurodiverse cognitive architectures in human will lead to wildly different systemic behaviours and representations. Different functional solutions, architecture of information processing, organs and role will cascade into different behavioural and cognitive Governance outcomes.

Adopting a principle akin to that in the *bit parity* example introduced in section 3.1 — showcasing emergent phenomena in which constraint only present at higher level assemblies affect the state of units from which such constraints are irreducible — externalities can be seen as such constraint that not only attests the emergent nature of governance cognition, but it also illustrates the sorts of cognitive biases higher level cognition can incur. In a parity bit system, individual bits possess no property corresponding to “parity”, given that this property exists only at the system level, as a constraint on collective configuration. The higher-level constraint family projects downward as conditioning $P(X|Y)$ that reweights the lower-level distribution, with the parity constraint registering at the bit level as a boundary condition that no individual bit’s marginal distribution can express. Similarly, externalities represent constraints that exist only at the governance level, invisible from the perspective of any individual market participant yet causally efficacious at system scale. This is downward causation through conditioning of lower-level distributions on higher-level states³²⁸. In a parity bit system — where, e.g. three bits must together yield an odd number of ON states — the restraint is imposed by an overarching parity rule applied to the entire bit string other than by any local interactions among the bits. While any two bits can be freely chosen, the rule forces the third bit to assume a value that ensures the whole

³²⁶This and all further quotes in this paragraph are taken from (Harari, 2024)

³²⁷(Harari, 2024)

³²⁸That section 1.3.4 formalized

system complies with the odd-parity condition. This global rule — often used in error-checking schemes — is not evident from observing any single bit or subset of bits, thereby illustrating a strong emergent property, the coordinated behavior arises only at the system level, not from the properties of the individual bits (Bar-Yam, 2004). Although out of place within the structure of this thesis — given that it is mainly in the synthesis chapter that I will be exploring the merging of political and cognitive science concepts within a common informational functional framework — it is important for the current idea at hand to provide a concrete example. In political science and economics, *Externalities* are costs or benefits not reflected in market prices, resulting in over-production of harms or under-production of benefits (in relation to optimum), with potential spillover effects leading to market or policy failures (Paniagua, 2023).

This can be informationally understood as an inability to perceive and integrate certain costs, benefits, or risks that fall outside the established metrics, variables, weights and parameters enabled by the governance protocol. In this sense, the Governance systems would suffer from a sort of *system blindness* bias due to not having the mechanisms whereby to represent and sustain a generative model which accommodates such features, regardless of individual humans being able to represent them. Consider a given country navigating a GDP problem-space (more on Governance problem-spaces, generative world model, search and exploration on the synthesis chapter), while lacking pollution perception-enabling protocols, institutions or features to represent it. In this real world example, which has been running for decades, the generative model of the governance system was constituted by tangible market and social representations which failed to represent environmental costs, thereby incurring in decision that showcase biases skewed against health variables, sustainability and long-term economic output, economic wellbeing and social well being. The consequences of this is a set of constraints which affect multifaceted degrees of freedom on individual humans (from internal health to external section of habitational zones), underpinned by mechanics irreducible to a single human state (as per checking a single bit state in bit parity restrains), which become tangible only at the governance system level, as biases in cognitive processes.

Such cognitive processing shortcoming is also present at the level of the individual human, and represents inevitable manifestations of cognitive processes leveraging types of information over other or running types of cognitive architectures over others. When a given individual is caught by change blindness³²⁹ or attentional bias³³⁰, the value of attentional states narrowing the set of variables relevant at the start of a cognitive process will make it so that information which otherwise be relevant is missed, and therefore incur in a bias. On the “architecture of cognition” front, when an individual human is caught in the shortcoming of Availability Heuristic or Bounded Rationality and Temporal Myopia, the structure or flow of information processing — although relevantly appropriate in circumstances — becomes a liability in certain settings. Regardless, none of such biases at the human scale is perceivable at the scale of the neuron. Likewise, higher level governance cognitive system

³²⁹ As per Simons and Rensink (2005), our attentional system — by narrowing the range of variables processed at any given moment — optimises cognitive efficiency in a trade-off with broader scene analysis, e.g., in the flicker paradigm, observers often fail to detect that an actor’s identity has been switched during a brief interruption — demonstrating how selective attention leads to the omission of relevant information

³³⁰ An Individual’s initial cognitive state may become overly focused on emotionally salient stimuli, excluding other less salient but nevertheless relevant stimuli as variables from processing. For example, in a dot-probe task, participants with threat bias fixate on threatening images at the expense of neutral cues — the experiment yielded significant differences in reaction times and eye-tracking patterns — with the selective focus on threat-related information assessing meaning faster than that of critical contextual data that was neutral, thus omitted from evaluation, showcasing a bias in decisioning operations with emotional weighting (Bar-Haim et al., 2007).

processes, pursuing system-level goals, aren't captured within the scope of the single individual human. This is due to the individual human becoming a functional agent within a Governance Cognitive System bureaucracy of agents, with own higher level cognitive states.

The preceding analysis carries direct implications for the constitutional design problem introduced in section 4.1, in that constitutional design is not merely institutional engineering, but the design of a cognitive architecture. Given that the Hard-Protocols articulated architecture dictates the information processing structure of the Governance System, it can be explicitly designed so as to accommodate functional modalities critical to enhance cognition across systems, present at most cognitive systems but (at least not optimised and explicitly present) absent in Governance ones.

Thus *constitutional* design is cognitive architecture design, and the variations between governance systems (centralised versus distributed, which variables rendered, which modules connected) are parameters of cognitive systems that determine their characteristic capacities and pathologies. That, monist consequences of informational architecture rather political or ideological, beyond the ought and merely within the realm of the *is*.

4.3 Part the whole

While section 3.2.4 established how characteristics from individual level humans binded by cultural scaffolding in *assembly index* route shaped the cognitive glue by which higher-level Markov blankets emerged, we needed the additional context of sections 4.1 and 4.1 to understand how the choices, biases, information architecture restraints and goals of system level Governance Cognitive Systems — in turn — exert downward causation influences back into the individual human level agents.

Recapitulating from section 3.2.4³³¹, the cultural pruning of attentional state-space that from the agent perspective corresponds to individually inaccessible cognitive states, can be understood from the higher level system perspective (e.g. Governance system enabled by the cultural scaffolding) as a type of bowtie compression mechanism (2.5), instantiating at the information processing level the compressed cognitive-states (out of all the possible state-spaces) at which range representations compatible with the cognitive scaffolding are viable. For example, in honour cultures (Nisbett & Cohen, 1996), attentional resources are systematically channeled toward reputation-relevant stimuli — such as honour status, slights, hierarchical positioning, loyalty, reputation — while cognitive states involving 'letting insults pass' or 'delegating grievance resolution to impersonal institutions' become attentionally inaccessible or require effortful override. This compression enables the coordination architectures characteristic of honour-based governance, such clan loyalty structures, blood feud resolution protocols, and patronage networks that presuppose agents whose attentional apparatus automatically tracks honour-relevant variables. By contrast, lawyerly societies which lean into institutions for conflict resolution, render *honour variables* attentionally costly, by defaulting into *legal remedies* with different cognitive attractors premised on impersonal rule-application. Following the same principles in which p53 compresses diverse cellular stressors into a single regulatory bottleneck that cells interpret contextually (section 2.5), cultural scaffolding *compresses* the vast space of possible human

³³¹ See 3.2.4 on cultural pruning of attentional state-space

attentional configurations into narrow bands that agents interpret through their local generative models trained on said attentional range, compressed into the spectrum that renders them legible to, and coordinable by, the culture it enables. Hutchins's (1995a) framing of culture not as a static repository of artifacts but as a dynamic cognitive process — a societally maintained cognitive function irreducible to any single human-level entity — anticipates this picture, with cultural scaffolding bearing exactly the parity-bit signature of section 3.1, with properties present at the collective level (the very meaning of an honour-relevant cue, the legibility of a legal remedy) cannot be read off any individual's marginal distribution and exist only as features of the joint distribution diagnosed by positive total correlation (section 1.1.3). At the higher, beyond human system level, the Governance systems operate on the compressed representations, which is, in this thesis' perspective, why cross-cultural governance transplantation so frequently fails, as the receiving population's cognitive compression is incompatible with the institutional architecture being imported, producing the interpretive failures analogous to the organ rejection dynamics discussed in section 2.5.

Cognitive psychology literature documents analogous architectural trade-offs operating at the individual neural level. As Tversky and Kahneman (1973) explain, our cognitive architecture is designed to rapidly retrieve instances from memory, a process that aids decision-making by quickly offering relevant examples from which to base and inform decision-making. However, when individuals rely on an availability heuristic, this very mechanism can be misleading, i.e., in one study, participants judged words beginning with the letter K as more frequent than words with K in the third position simply because the former were easier to recall, despite objective counts showing the opposite. This example illustrates how an information processing flow optimised for efficiency becomes a liability when the ease of recall replaces accurate statistical evaluation. Worthy et al. (2012) demonstrated that — while our cognitive architecture is typically adept at processing sequential, immediate feedback — under high working memory load, it becomes biased toward short-term gains. This is a biased liability in settings that demand long-term planning, i.e., in decision-making tasks, participants engaged in a concurrent numerical Stroop task (so, experiencing reduced working memory capacity) disproportionately favored option with the highest immediate reward, even when selecting alternative options would have yielded greater cumulative benefits (over time). That is, the efficiency gains that enable rapid decision-making under typical conditions, become liabilities under atypical loads, as the bowtie compression mechanisms of section 2.5 necessarily discard information that may prove contextually critical.

Akin to how cell's individual competence and novelty (in Xenobots and Anthrobots) is suppressed when pruned by the state-space compatible with a multicellular organism³³², so does human's reliance on Culture's *cognitive niche construction* has a dual effect on adaptability. Think of memory and the written language. Individual human cognitive capacities to store and recall memories — which, prior to written language, would vary personally from song encoding, to repetition, rituals, and higher reliance on internal neural structures of memorisation alone — become with language standardised, conventional and abstractable from any one given individual. While each human post literacy, by offloading memorisation to calendars, notes and diaries, will rely less on internal neural structures.

³³²As discussed in section 2.2, cells freed from bioelectric coordination express wider behavioral repertoires but lose access to the morphological possibility spaces that coordinated collectives can navigate. The constraint-enablement signature is invariant across substrates, what changes is only the medium through which constraints propagate (bioelectric signals, linguistic conventions, legal codes) and the characteristic timescales over which integration and constraint operate.

offloading initiates evolutionary feedback dynamics with long-term consequences for human cognitive architecture. This dynamic reflects a coevolutionary pattern, once a cognitive innovation is offloaded to culture, there is less evolutionary pressure for individuals to do it all mentally, which in turn further "locks in" the reliance on cultural support, enforcing a loop. This coevolutionary dynamic was formalised in evolutionary biology, "*In recent years, a number of evolutionary theorists have spoken well of a turn-of-the-twentieth-century idea that since the nineteen fifties has gone by the name of "the Baldwin effect" (Hinton and Nowlan 1987; Dennett 1995; Deacon 1997). The general thrust of the idea is to urge that, under some conditions, learned behaviors can affect the direction and rate of evolutionary change by natural selection. In such cases, cultural inheritance of a learned behavior across an indefinite number of generations creates a "breathing space" (Godfrey-Smith, chap. 3, this volume) in which inherited factors favorable to the adaptive behavior in question that either already exist, happen to crop up, or can be stimulated by the change in question—there is some dispute about this—will move along the channel already cut by culture, thereby converting learned behaviors into genetic adaptations or, alternatively, supporting learned behaviors by related genetic adaptations. In either case, natural selection will have ratified evolutionary vectors that learning began*" (Weber & Depew, 2003, p.3). Baldwinian effects thereby compromise individuals' independent adaptive range³³³, steering biological development toward roles tied to cultural practices. This is expected as the type of component specialisation at lower levels when articulated within a higher-level system³³⁴, as I've illustrated with 2.2's analysis of sieve tube cells sacrificing organelles for transport function. As more tasks get culturally uploaded — abstracted from unique individual implementation while standardised so as to become accessible across the collective — the abstraction bottleneck (section 3.2.2) that enables constitutional assemblies simultaneously creates the selective pressures that genetically assimilate the cognitive prerequisites for operating within those assemblies, thereby engineering a human substrate increasingly fitted to the governance systems it constitutes, which in turn enables further constitutional sophistication demanding further human specialisation. Albeit not explicitly framed in this manner, Hutchins already rejected a static definition of culture — framing it as a dynamic, adaptive process³³⁵ — and as a cognitive process in itself³³⁶. Understanding culture as a *de facto* cognitive process (and as artefacts with assembly index) has deep implications both in human evolutionary Baldwinian speciation, as more tasks get *culturally uploaded*, thus abstracted from unique individual implementation while standardised so as to become accessible.

These are, as the concept of downward causation introduced in the Preamble first foreshadowed, influences of the whole on its constituent parts, what section 1.3.4 reread as the conditioning of lower-level distributions on higher-level states, where higher-level constraint families project downward as $P(X|Y)$, reweighting lower-level state probabilities without recruiting any new physics. Cultural pruning of attentional state-space (section 3.2.4) and the Baldwinian assimilation just traced make this tangible at the cognitive level,

³³³ This dynamic reflects a gene-culture coevolutionary pattern: once a cognitive innovation is offloaded to culture, there is less evolutionary pressure for individuals to do it all mentally, which in turn further "locks in" the reliance on cultural support, enforcing a loop.

³³⁴ In the sense that Higher level evolutionary pressures lead to lower lower level changes. Generally speaking, the fact that bird culture permeates mating calls as a fitness mechanism has lower level downstream effects on changing cellular structures to fit mating call functions.

³³⁵ "Culture is not any collection of things, whether tangible or abstract. Rather, it is a process. It is a human cognitive process that takes place both inside and outside the minds of people." (Hutchins, 1995b)

³³⁶ "I am proposing an integrated view of human cognition in which a major component of culture is a cognitive process (...) and cognition is a cultural process." (Hutchins, 1995b)

pathology makes it tangible at the somatic one. Take GDP growth as a target variable within the cognitive-light-cone of most contemporary nation-states. The imperative to maximise productivity does not float at higher level alone — it is only achievable through the restriction of alphabets of states within the nation's lower level Markov blankets, cascading — through Hard Protocols in the form of labour codes, working-time directives, agricultural subsidies, educational throughput, urban planning frameworks and property-market structures — into the individual human substrate concentration of available states. In the pursue of higher levels goals, such as higher GDP, what emerges with statistical regularity is sleep deprivation (Hafner et al., 2017; Walker, 2017), the association between long working hours and depression, anxiety, ischemic heart disease and stroke (Pega et al., 2021; Virtanen et al., 2018; Kivimäki et al., 2015), loneliness (formally classified as a public-health emergency) (Murthy, 2023; Holt-Lunstad et al., 2015; Cacioppo & Cacioppo, 2018), the inflation of ADHD diagnosis as productivity demands and pharmaceutical-policy regimes interact (Conrad & Bergey, 2014; Hinshaw & Scheffler, 2014), the deaths-of-despair epidemic that Case and Deaton (2020) traced to the deindustrialisation logic of late-twentieth-century capital allocation, and the blunted neurocognitive development of children with inadequate green spaces (Reuben et al., 2019; Dadvand et al., 2015; Bellinger et al., 2016). What clinics treat as an individual illness is, under this frame when taken into account the correct level of analysis, the very consequence of human embodiment into the path-reweighting of section 1.3.4.

Like different organismal solutions lead to different cellular stressors and more or less common failure modes, different governance configurations select for different attractors and therefore for different pathology profiles. A militaristic regime selects for one set of stress phenotypes (combat trauma, conscription-cycle morbidity, hypervigilance disorders) while a theocratic one lead to another (scrupulosity, ritual-compliance anxieties, sexual-shame syndromes), and a growth-maximising one for those mentioned above. Once one accepts humans as the operational substrate of governance assemblies, the health of the substrate is no longer separable from the cognitive operations of the system that conditions it, in the same way that the morphology of a sieve tube cell is no longer separable from the plant-wide sugar transport function it serves. An early articulation of this instrumental relation appears, inadvertently in regards to its implications, in Cyert and March who, while developing their behavioural theory of the firm, observed that *“just as adaptations at the individual level depend upon phenomena of the human physiology, organizational adaptation uses individual members of the organization as instruments”* (1963, Chapter 6). The framing of their remark stays at the firm scale while not developing its full implications — they were diagnosing how firms adapt, not theorising governance assemblies — but the structural observation translates. Organisational adaptation operates *through*, not *despite*, the embodied physiology of its constituent agents. What the firm uses *as instrument*, the cell uses *as morphology*. What the cell uses *as morphology*, the governance assembly uses *as constituency*. The categorical separation between agent and assembly dissolves once the instrumentation relationship is taken seriously at every scale on which it actually operates, with the dualist temptation to keep them analytically apart being the inheritance section 5.2 diagnoses as a part of Political Science's blind spot. What differs between cellular and human cases is only the interpretive bandwidth (section 2.6) available to the lower-level agents. Humans can up to a point model the system that conditions them, even if they cannot exit it, while cells cannot. While writing this chapter, I could imagine a magical realism short-story about a *doctor-cell* treating a patient cell on my body for oxidative stress, without ever being aware that the

organism both inhabit is running on three hours of sleep and instant noodles to finish this thesis.

Evidence across multiple domains supports this downward causation framework. For instance, as Kirkbride et al. (2024) demonstrate, in dual-earner governance environments – where institutional and legal structures *explicitly promote* economic parity by enforcing or enabling both men and women to contribute to household income – the gradient of mental health inequality across genders is diminished³³⁷. In this case, the mental states of individuals are not only shaped by the intimate experience of inequality or lack thereof, but by systemic, structural goals embedded within the governance system's operational logic. The redistribution of childcare labour, workplace inclusivity, and domestic expectation burdens are not simply *cultural shifts* but the embodiment of governance-level cognitive objectives seeking to optimise workforce participation and social cohesion. It bears emphasising that I am being descriptive while offering no prescriptive judgement regarding the labour market inclusion of women³³⁸. What is evident is that both Hard-Protocols and Cultural-Scaffolding directly impact variables with systematic repercussions in biology and psychology, with any move carrying trade-offs across variables that can be empirically measured against a hierarchy of values selected by the cohort. The framework simply makes these trade-offs visible as features of the multi-level constraint architecture instead of collapsing them into a single normative axis.

This embodiment is also visible in housing, with Rahman and Steeb (2024) showing that renters – who inhabit a structurally more precarious relation to both capital and community – face measurably higher odds of mental health disorders compared to homeowners. It is then easy to see how higher level protocols encoding zoning, property law, borrowing, and many such other problem-spaces navigated by governance cognitive systems directly influence lower level psychological state space at scale. Air pollution – which has been shown to predict depressive and psychotic symptoms – too must be understood as the material downstream of urban zoning, transportation policies, and industrial prioritisations and such, variables themselves reflective of governance systems navigating higher cognitive-light-cone goals. Obesity and attention-related disorders – cast as lifestyle diseases or individual psychiatric dysfunctions – seem to reveal themselves as expressions of governance dynamics too, as showcased by individuals with obesity presenting comorbid mental disorders such as ADHD, anxiety or depression, articulating how ADHD is entangled with lifestyle and societal factors originating not *purely* at the individual level but emerging from a systematic architecture in which food subsidies – resulting from higher-level governance cognitive-light-cone pursuits – distort nutritional availability and urbanisation limits – likewise resulting from higher-level governance problem-spaces such as commute efficiency, tax base concentration, or industrial clustering – hinder physical play, engagement and mobility.

From this integrated, multiscale agentic higher-level assembly perspective, one can interpret news such as “77 percent of American youth can't qualify for military service” as an embodiment consequence of higher-level system behaviour which leads to a compromised defense system. That is, not a dysfunction to be located in the lower-level units alone but a reflection of the Governance Cognitive System's behaviour and *trajectory navigating* the

³³⁷ “in countries with a dual-earner model, where employment, wage earning, and domestic and childcare tasks are shared more equitably between men and women, gender inequality in mental health risks appears to be smaller” (Kirkbride et al., 2024)

³³⁸ Which has done wonders for female freedom, autonomy and self-determination, even while carrying further implications across variables such as natality, wages and dating culture

myriad problem-spaces across its cognitive-lightcone, and its impact on the embodiment of its lower level substrate as a consequence. This brings forth an important dynamic not yet addressed – yet inescapable once humans are understood no longer as external but as the embodiment and constitutive substrate of governance systems – derived from the Red Queen Effect. *“Leigh Van Valen was an American evolutionary biologist who made major contributions to evolutionary theory. He is particularly remembered for his groundbreaking paper “A New Evolutionary Law” (...) where he provided evidence from fossil record data that the probability of extinction within any group remains essentially constant through time. In order to explain such an unexpected result, Van Valen formulated a very influential idea that he dubbed the “Red Queen hypothesis.” It states that the constant decay must be a consequence of evolutionary interactions among connected species within ecological networks. In Van Valen’s picture, species do not merely evolve: they also coevolve with other species. (...). Van Valen’s law provided the first complex systems theory of coevolutionary dynamics and inspired a whole range of theoretical and experimental developments from very diverse fields, percolating far beyond its original formulation. Red Queen arms races are nowadays considered a widespread feature of complex adaptive systems”* (Solé, 2022).

In the same manner *Hard Protocol’s* higher-level pressures are pathologised in lower-level human agents, the same is true for *positive* outcomes — i.e. when higher-level Governance Cognitive Systems’ downward causation shape the alphabet of attainable microstates of lower-level agents through attractor and motivational dynamics that capture their behaviour — exerting pressures within human agents into continuous adjustments that select for system-compatibility. Lower level adaptability, that is, the socio-cognitive compliance to the direction of higher-level *Hard Protocol* change, as Governance Systems respond to their problem-spaces by engaging in continual structural transformation — be these manifested in labour requirements, economic policies, normative prescriptions and so on — that lower-level agents must assimilate, result in intra-lower-level performance pressures towards system level fit. The trait’s fit is established not because it is intrinsically and autonomously *better*, but because it aligns with higher-level pressures.

For example, consider a System level pressure around educational credentialism, through the lens of a *Red Queen dynamic*, within *Hard Protocols* structured around enforcing meritocratic allocation of roles and resources. While also factoring in governance’s assembly indexes increasingly abstracting human roles into standardised functional categories, adding requisites for specific qualifications trigger human agents to engage in an escalating *arms race* of credential acquisition. When a bachelor’s degree becomes the minimum requirement for administrative positions within governance institutions, individual agents respond by pursuing higher credentials, which subsequently shifts the competitive landscape such that master’s degrees become a new minimum, leading to requirements that previously required only basic literacy. This mechanic remains the same across archetypical *Hard Protocol* incentives depending on the type of Governance Cognitive System, from militaristic societies to theocratic ones as the System’s cognitive-light-cone encodes different success variables as it navigates its problem-spaces. Just like biological Red Queen scenarios, humans within governance systems cannot exit the competitive landscape without severe fitness penalties (unemployment, social exclusion, economic precarity, governmental coercion...), which leads directly to the consideration of the Baldwin Effect within higher-level governance assemblies, with learned adaptations becoming genetically assimilated over successive generations subjected to the long-run selective pressures that favour the embodiment of traits aligned with the overall higher-level system’s fit.

This conjunction of downward causation pressures and Red Queen competitive dynamics creates the conditions for Baldwinian genetic assimilation at the governance scale given enough evolutionary time. Given that human agents within higher level governance cognitive systems face fitness penalties for misalignment, while also favouring learned traits aligned with System's pressures through the Red Queen Dynamic, its therefore established the fundamental mechanics to rationally argue that — to a bigger or lesser extend and adjusted to the very short evolutionary period of Constitutions in assembly history — over time these selection pressures cascade from Constitutional Code to DNA code.

However, as the Baldwin Effect would enable increased system-level fit, it would also incur a deep cost, as the human level make-up — subject to evolutionary pressures steaming from the two system-level mechanisms above — becomes increasingly tied to a higher level definition of fit, which optimises for functional and abstractable roles, while also reducing individual demand for capacities which are outsourced throughout the system. For better or worse, the traits for which a Governance Cognitive System selects (and some traits are foundational to the assembly history and index of having a *constitution*, such as abstract reasoning, conformity to procedural protocols, etc) are effectively engineering their human substrate, and by influencing their embodiments are changing themselves. This is a fundamental principle which is expected of organisms going from lower to higher level, such as of living organisms going from single cell to multicellular system.

But the relationship is bidirectional. Section 2.6 established the hard protocol / dynamic drive architecture, where mechanistic constraints establish possibility spaces while adaptive interpretation determines instantiation. The manifestation of any constitutional code depends on the dynamic drives of the human agents who interpret and enact it. The relationship between Dynamic Drives and Hard Protocols, can be manifested in a myriad or bidirectional manners. To illustrate the phenotype effect of *Dynamic Drives* over *Hard Protocols* take two examples from Portugal, the jurisdiction I am most familiar with³³⁹, to see cases in which agentic and cultural-scaffolding aligned motivations made it so that lower level human agents didn't run or enforce *Hard Protocols* at scale, therefore having non-manifested or non-expressed higher-level code: despite (prior to the reform of 2007) Portugal having had criminal statutes against abortion, the legal system did not enforce them systematically³⁴⁰; and the same is true for drug consumption (prior to the 2001 reform), in which criminal drug consumption laws, although exiting, weren't actively enforced³⁴¹.

These are just two narrow examples in which lower level agents affect higher level embodiment, amidst a constellation of all the other ways in which higher level Governance System's are emergent from Constitutional code being dynamically applied and interpreted — rather than legal code simply encoding a governance system in a blueprint-like way — thus being a co-constructed, enacted, interpreted and dynamically solved emergent property of lower level human agencies navigating dynamically problem-spaces in interpretation, cost-benefit, cultural-scaffolding motivation, goal pursuit and so on, in a multi-scale, multi-competent, overlapping of cognitive-lightcones. Yet, in the same manner in which lower

³³⁹ Having graduated in Law in Lisbon.

³⁴⁰ "As Alves, Santos, Barradas and Duarte, (2009) stated: "Indeed, between 1998 and 2004 were recorded, at the Ministry of Justice, 223 crimes of abortion which resulted in 34 cases, 43 offenders and 18 convictions" (Figueiredo & Silva, 2014, p. 6); "In 2007, given the estimated 23,000 illegal abortions in Portugal annually, and the over 10,000 Portuguese women hospitalized because of these clandestine abortions, pro-choice groups decided to focus on the human suffering" (Manuel & Tollefsen, 2008). Triangulating both these quotes, we can see that only 0.078% of illicit abortions converted into convictions, meaning that the law was virtually non-enforced.

³⁴¹ "It has been estimated that one percent of the Portuguese people were addicted to heroin (...) Notably, however, individuals punished with incarceration for possession or use of illicit drugs did not comprise a significant portion of Portugal's prison population at the end of each year from 1993 to 2000 (...) the data suggest that drug consumers were rarely punished (...)" (Kanoho et al., 2017, p.8)

agents affect the higher lever embodiment of *Governance Cognitive Systems*, downward causation effects of *Hard-Protocols* also have a generalised effect on the embodiment of lower, human level agents. This parallels section 4.1.2's view of constitutional code and DNA, neither operating as a deterministic blueprint. As DNA is dynamically applied and interpreted through multi-scale cellular agencies navigating problem-spaces in selection, transcription, and translation, so too is constitutional code dynamically applied and interpreted through multi-scale human agencies navigating problem-spaces in legal interpretation, enforcement discretion, and cultural motivation. The relationship is bidirectional, constitutional code shapes DNA code through the Baldwin Effect, while cultural-scaffolding and individual dynamic drives shape the phenotypic expression of constitutional code.

Thus, the five mechanisms³⁴² traced above converge on a fundamental reconceptualisation of the governance-human relationship. The monist framework sustained throughout this thesis reaches its culmination here. On the opposite spectrum of isolating disciplines and approaches, one ought to regard Governance Cognitive Systems as entities in assembly space like any others, as part of physics, biology and ecology, recursively shaping its own substrate. It should no longer be regarded as *external* to the human than the ribosome is to the gene, and therefore, the notion of policy-making reconceived not as exogenous interventions but as morphogenetic. Like organisms to cells, Governance Cognitive Systems are informational protocols that shape the viability of entire modes of being. Interiorising this leaves us with one of the most striking conclusions afforded by this monist approach: that Governance Cognitive Structures are no different than any other cognitive system, that, from a first-principles perspective, there is no functional difference — nor in assembly, role or process of maintaining integrity — between a cell's organelle, a human's organ, and a Constitution's organ.

4.4 Recap: Deriving Governance from first principles

This section aims at shooting the multiple Chekhov's guns showcased throughout this thesis, distilling the incremental arguments chained throughout into a single core stream.

The Problem

Cognitive Science and Political Science occupy different epistemological realms, with different terminologies, paradigms, and units of analysis. Making progress in both cognitive science and political science – so as to understand cognition at scale and be able to model, understand and design better policies, organs and political systems – requires a common analytical substrate capable of synthesising these two realms into a common floor from which first principles can be derived and applied across fields (Preamble).

The guiding constraint

The synthesising substrate must be (i) anchored in first principles and compatible with a monist, physicalist, epistemology; (ii) capable of operating across levels and scales without conceptual rupture; and (iii) robust against the ontological critique of reductionism (that higher-level phenomena may not have explanations at the ultimate level) plus the

³⁴² cultural constraint-enablement; downward causation from Hard Protocols; Red Queen competitive dynamics; Baldwinian genetic assimilation; and bidirectional Dynamic Drive effects,

downward causation critique (that higher-level phenomena exert causal effects on lower levels) (Preamble).

Step 1 — First Principle:

Information as the measurable departure from maximum entropy

Any system occupies one configuration drawn from a probability space — a triple of sample space, collection of events, and probability measure (section 1.1.1). The sample space fixes the alphabet of microstates the system could occupy; the probability measure assigns each event a number between zero and one capturing its likelihood. Information is the measurable departure of a joint distribution from its maximum-entropy reference. Section 1.1.2 disambiguated three complementary registers of this single concept: negentropy (the constraint content of a single distribution, measured against the uniform reference), mutual information (the coupling between two random variables, including across levels of description), and total correlation (the coupling among the components of a system at one level, measured against the product of their marginals). Together these capture both the temporal acquisition of knowledge — when a distribution concentrates upon the arrival of new data — and the structural instantiation of order — when a distribution concentrates because of organisational relationships among components. Both modes are entropy reductions; both are gains in information. The framework is substrate-independent, requiring nothing beyond a specifiable probability space (Section 1.1).

Step 2 — Filter:

Systemhood as positive total correlation

Three independent dice yield 216 equiprobable joint outcomes; knowing one die's state tells nothing about another's. Cellular organelles, by contrast, exhibit functional dependencies — ribosomes requiring mitochondrial ATP, membranes requiring enzymatic maintenance — that constrain the joint distribution to occupy only a subset of the theoretically possible joint configurations compatible with cellular integrity. Section 1.1.3 formalised this separation: a multi-component level constitutes a system when its total correlation is positive — when the joint distribution of its components is concentrated on a strictly smaller alphabet than the Cartesian product of the component alphabets would allow. Total correlation captures Weiss's substantive intuition (that systems exhibit overall invariance despite component fluctuations) without the failures of variance-based formulations: it registers all forms of statistical dependence, positive or negative, linear or non-linear. The dice collection has zero total correlation and therefore is not a system; the cell has positive total correlation across its organelles and is. A system, then, is an informational entity whose components are mutually less free than they would be in isolation, whose joint distribution sits below the maximum-entropy reference (Section 1.1.3).

Step 3 — Cross-level structure:

Coarse-graining and the recursive nesting of levels

The same probability-space apparatus supports multiple nested levels of description. A coarse-graining is a function that partitions the fine-grained microstate alphabet into equivalence classes — groups of microstates indistinguishable at the coarser level — and

replaces a distribution over a larger alphabet with a distribution over a smaller one. The higher-level probability follows by pushforward: each macrostate's probability is the sum of fine-grained probabilities over its equivalence class. Levels nest recursively, with each higher level being one function-composition further from the base than the level below it, and the entire hierarchical structure being fully determined by the base distribution and the chain of coarse-grainings that connects them. Macrostate descriptions are not impoverished versions of microstate descriptions but new variables germane to the higher level. The total fine-grained entropy decomposes into the entropy of the macrostate plus the average within-macrostate residual entropy. This is the formal grounding for GST's hierarchical and nested picture: each level is a different way of labelling the alphabet of the same underlying probability space (Section 1.1.4).

Step 4 — Bottleneck:

Assembly memory (systems as historically accumulated constraints)

Assembly Theory reconceptualises objects as entities defined by their "formation histories". The assembly index measures the minimal recursive operations required to construct an object from elementary building blocks. Copy number counts identical instances. High copy numbers at high assembly indices are combinatorially impossible without selection acting on historically realised construction paths. Successful constraint configurations persist as reusable memory, i.e. assembly bottlenecks exhibiting convergent constraint satisfaction, reusability and informational compression. The peptide bond satisfies thermodynamic, geometric, and functional constraints simultaneously and, once assembled, becomes a reusable step incorporated into the pathways of a myriad of protein structures. Without such bottlenecks, systems would require impossibly exhaustive combinatorial exploration. With them, each constraint step channels the next (Section 1.2).

Step 5 — Phase Transition:

Singularity-nodes and selection as path-reweighting

Within the hierarchical architecture of Step 3, certain configurations are not smooth points of the constraint family that produced them. Section 1.3 introduced the formal apparatus for identifying these. A constraint family is a one-parameter family of distributions over the same components, indexed by some parameter governing the constraints. As that parameter varies, the free energy of the constrained ensemble (the trade-off between energy minimisation and entropy maximisation under the prevailing constraints) usually depends smoothly on the parameter. At certain critical parameter values, however, this dependence becomes non-analytic — derivatives diverge, new macrostates become stable, the structure of the configuration space the system effectively explores reorganises. These are phase transitions. The information-bottleneck Lagrangian provides a parameter-free analogue applicable to any joint distribution with a relevance variable: bifurcations of the optimal representation under the compression-relevance trade-off identify the same kind of qualitative transition without requiring a physical free energy. A singularity-node is the configuration in configuration space where one of these events occurs — a non-analyticity of the free energy of a constraint family (section 1.3.2) or a bifurcation of the optimal representation in an information-bottleneck Lagrangian. Singularity-nodes are where new levels of description consolidate into stable post-transition phases.

Epistemological stepping stone:

Information as the measurable departure of a joint distribution from its maximum-entropy reference, on whatever probability space the level of inquiry requires, with viability (which configurations get reused), assembly bottlenecks (where in the assembly graph the reuse concentrates), and singularity-nodes (where in the parameter space the consolidation that makes a level happens) as its highest expressions, provides the reductionist substrate. Systems emerge and maintain themselves by accumulating constraint-cascade memory along assembly paths, consolidating new levels at singularity-nodes of their constraint families, and projecting downward as conditioning that reweights lower-level configurations

Step 6 — Bottleneck:

The non-equilibrium threshold, from passive to active constraints

Passive constraints — chemical & physical bonds, geometric compatibilities... — maintain assemblies in equilibrium. A crystal grows through passive constraint satisfaction requiring no ongoing energy expenditure. As assembly indexes increase, however, passive mechanisms prove insufficient against entropic decay. This enforces an absolute bottleneck, non-equilibrium systems either drift back to equilibrium or develop minimal homeostatic solutions to persist. Only configurations capable of transitioning from passive to active constraints, that is, dynamic processes that monitor system state and execute corrective actions, survive beyond this threshold. What becomes possible after crossing it is an entirely new regime of assembly exploration, requiring both preferences, priors or bias, i.e., non-uniform preferential distributions for certain states that have systematically higher probability of encoding or promoting configurations which maintain integrity; and a way to check and steer current states towards those preferred ones (Section 2.1).

Step 7 — Phase Transition:

Cognition as active integrity maintenance

From preferences, Active Inference formalises how non-equilibrium systems persist by minimising prediction error through coordinated perception-action cycles. The system maintains a generative model encoding predictions about environmental states and acts to reduce discrepancy between predicted and observed outcomes. Cognition is not something systems achieve at a threshold of sophistication, it is rather what enables non-equilibrium systems to exist at all. Information gain now operates through two complementary registers: the static register from Steps 1–6 (negentropy, total correlation, conditioning, path-reweighting) describing how constraints consolidate stable post-singularity phases, and the dynamic register, where Active Inference prunes generative models incompatible with homeostasis through prediction-action loops. The free energy minimised dynamically by an active-inference agent is the same kind of object that section 1.3.2 used statically to locate singularity-nodes: agents act so as to keep their state distributions concentrated on the small alphabet of configurations compatible with their integrity, mirroring at the dynamic level the alphabet compression that constraint families effect at the static one (Sections 2.1, 1.1–1.3).

Step 8 — Framing:

Cells as nested cognitive agents

Cells exemplify this architecture, as historically contingent agents navigating assembly problem-spaces along two pruning axes: constraining inherited lower-level state-space; and refining generative models of environmental interaction. The nested character of this architecture is visible across scales simultaneously. Within the cell, mitochondria operate as autonomous problem-solvers, integrating calcium and stress signals, deciding between ATP production, mitophagy, or apoptosis, and triggering immune responses by releasing mtDNA when persistently damaged, maintaining their own constraint-satisfaction procedures while serving cellular-level goals (Section 2.2).

At the cellular level, planaria exposed to barium, an evolutionarily novel stressor never encountered in their evolutionary history, regenerate resistant heads by navigating unfamiliar transcriptional space, demonstrating cellular agents solving problems their individual generative models never encoded (Section 2.4). Beyond the cell level, organism-order goals reshape cellular state-space, with sieve tube cells — for example — beginning as typical plant cells possessing nucleus, vacuole, and organelles, yet during differentiation losing these structures to become open transport pathways for plant-wide sugar distribution. This is downward causation operating exactly as Step 6 described: the plant-level constraint family conditions the cellular distribution, with the higher-level state setting parameter values under which cellular differentiation proceeds, channelling some configurations and pruning others. Each level retains problem-solving autonomy, while being simultaneously shaped by the level above and composed of autonomous agents below. The multiscale competency architecture that ensures collective persistence (Section 2.2).

Step 9 — Framing:

Cognitive glue and multi-agentic assembly

Agents at the beyond-equilibrium threshold, akin to simpler at-equilibrium subsystems before them, face the same coordination dynamics formalised through coarse-graining, conditioning, and total correlation in Steps 3 and 6. Individual cells, each maintaining autonomous homeostatic drives as established in Steps 8 to 9, face this coordination challenge. How can distributed agents whose generative models encompass only local states, align constraint-satisfaction procedures with system-level goals none of them individually represents?

At the cell level, bioelectric communication — ion channels, gap junctions, voltage gradients predating specialised neurons — provide a bridging medium. By creating conditional dependencies between cellular states, bioelectric signalling drives the total correlation among cellular states above zero — satisfying the systemhood criterion of Step 2 at a new scale — where knowing one cell's voltage state reduces uncertainty about its neighbours', concentrating the joint cellular distribution on a strictly smaller alphabet than the Cartesian product of the cellular alphabets would allow. Higher-level Markov blankets consolidate as statistical boundaries encompassing the collective, with sensory and active states belonging to no single agent. For example, picasso tadpoles whose scrambled organs navigate novel paths to correct morphology, or planarian fragments regenerating complete organisms from compressed bioelectric pattern memory, represent collective problem-solving capacities irreducible to any constituent cell's generative model. Any informational substrate that binds distributed agents into an integrated system, by establishing collective-level conditional dependencies exceeding individual capacities, is

what I've understood as cognitive glue.

Critically, cells freed from bioelectric cognitive glue (Xenobots, Anthrobots) spontaneously innovate novel structures, demonstrating both inherent cellular autonomy and the constraining role of the collective. Meaning that integration restricts individual repertoire while enabling collective possibility spaces, inaccessible to isolated agents. The same constraint-enablement signature observed at every prior scale, from electrons bound within atoms gaining access to molecular chemistry (constrained at the atomic level, enabling molecular-level assembly), to non-equilibrium systems gaining access to active inference by accepting the constraint of preferences at Step 7 (Sections 2.2, 2.4, 3.2.4).

Step 10 — First Principle:

The bounded scope of unidimensional agents in multidimensional problem spaces

Non-equilibrium agents maintain generative models and that multi-agent coordination creates collective capacities exceeding individual ones. Thus the necessary corollary follows that every generative model is bounded. The most distant set of events a system can measure and attempt to regulate in its goal-directed activity defines what Levin terms its cognitive light cone. This is the horizon of what the system can sense, predict, and act upon. A cell's cognitive light cone encompasses local chemical gradients and bioelectric states, while an organism's extends to spatially distant and temporally extended information.

The diversity of variables across cognitive light cone's levels, mean that three-dimensional navigation is just one problem space among a myriad of different problem spaces they encompass. If the variables a system must track to maintain integrity span dimensions beyond spatial containment — metabolic gradients, transcriptional states, social obligations, economic cascades — then no morphological boundary (membrane, skin...) can demarcate where the system's cognitive engagement begins and ends.

What delineates a system then the reach of its generative model, that is, the set of variables across which it can detect deviation from preferred states and act to correct it. System boundaries are therefore informational rather than morphological, defined by the scope of variables a system's generative model can integrate. When agents coordinate through cognitive glue (Step 8), their individual cones merge into a collective cone encompassing variables no constituent alone could model (Sections 2.4, 2.5).

Step 11 — Bottleneck:

Bureaucratic architectures as convergent coordination solutions

Bounded cognitive light cones (Step 9), within multi-agent assemblies enabled by cognitive glue (Step 8), generate a universal coordination problem across any nested system, regardless of level.

Sub-agents cannot model the full state of the assembly they participate in, yet they must align local constraint-satisfaction procedures with system-level integrity requirements that their individual cognitive-light-cones cannot encompass. The convergent solution (and an assembly bottleneck in itself, since multi-agent configurations that fail to consolidate it cannot advance toward higher-order assemblies) is what this thesis functionally defines as a *bureaucratic architectures*: bowtie compression-decompression cycles where multidimensional system states are compressed into low-dimensional *artifacts*, then decompress into context-appropriate operations by the receiving sub-system agents (and

vice versa in the reverse direction).

This architecture addresses a constellation of shared problems inherent to all distributed nested assemblies: 1) asymmetric information access across sub-agents occupying different regions of the system; 2) the need for storage and retrieval of information beyond any individual sub-agent's memory capacity; 3) categorisation frameworks that render compressed signals interpretable across heterogeneous receivers; 4) and the bilateral translation between system-level states and sub-agent-level actions, where local operations aggregate upward into collective function, while system-level requirements propagate downward as constraints on local possibility space.

Compressed bottleneck signals preserve sub-agent autonomy through controlled ambiguity. Piggypacking on active inference priors without dictating implementation, functioning as constraints on the probability landscape within which sub-agents continue exercising their own dynamic drives (Step 6). Compression necessarily creates blind spots, information discarded at the bottleneck cannot be completely recovered at the receiving end, yet without it, sub-agents would require full system-state access to coordinate, which bounded cones (Step 9) render impossible, and the collective Markov blanket (Step 8) could not maintain the statistical coherence required for further assembly.

All viable bureaucratic configurations therefore balance two complementary aspects derived from the hard-protocol / dynamic-drive architecture. *Mechanistic constraints*, that delimit the joint possibility space by specifying which sub-agent configurations remain compatible with system-level integrity, and *interpretive latitude* that enables adaptive decompression within those constraints, with the bandwidth available for interpretation scaling with sub-agents' cognitive light cone capacities (Step 9). Low-bandwidth protocols approach mechanistic rigidity where signals admit minimal contextual variation, while high-bandwidth protocols permit substantial situated judgment (the ratio constituting a design parameter any nested assembly navigates).

These principles appear identically across substrates and scales. p53 compresses diverse cellular stressors into a single regulatory hub whose activation state is contextually decompressed by downstream agents; the HPA axis compressing multidimensional threat information into corticotropin-releasing hormone that hepatocytes, immune cells, and neurons decompress according to their local functional requirements; planarian bioelectric memory compressing morphological instructions into voltage patterns that cellular fragments decompress into complete regeneration; and, as subsequent steps derive, documentary, linguistic, and legal systems performing structurally identical operations through different media at the beyond human scale (Sections 2.5, 2.6).

Step 12 — (re)Framing:

Distributed cognition and cognitive scaffolding at the human scale

The principles established at cellular scales apply with structural invariance at the human scale through different substrates. A Danish citizen's vastly superior capacity to recall events from a specific day ten years ago relative to an indigenous Melanesian without calendric concepts, reflects not hippocampal differences but participation in a cultural assembly of higher index, with calendars, digital records, banking statements functioning as externally maintained constraint-cascade memory. Failing to recognise this distributed architecture leads to misattributing system-level properties to individual agents — the same analytical error as attributing to individual cells morphological competencies that emerge from bioelectric coordination (Step 10), and structurally identical to the bit-parity error:

attributing to a single bit a property (parity) that exists only at the level of the joint distribution, which can be diagnosed by positive total correlation (Step 2) but cannot be read off any single component's marginal distribution.

The mechanism of cognitive scaffolding allows for embedded environmental, cultural and social structures to support cognitive operations that individual agents leverage, ranging from sponges exploiting ambient water currents to reduce active pumping, to humans using linguistic labels to compress perceptual categories. Each stabilised scaffold becomes an assembly bottleneck in its own right, as accumulated memory that subsequent users inherit without re-derivation, enabling navigation of problem spaces inaccessible to unscaffolded cognitions sharing an equal physiology. Without current, a sponge will restrain thriving in much the same way as a human will without language. However, the transition from passive environmental leveraging to active cultural construction marks a critical phase transition from exploiting existing passive structures to deliberately building structures that, by design, extend collective cognitive light cones (Step 11) beyond any individual's reach (Section 3.1).

Step 13 — Phase Transition:

Culture as assembly memory and the cultural Markov blanket

Through generational cognitive uploading — the active tailoring of shared physical, social, and symbolic environments to support cognitive operations at scale — individual scaffoldings (Step 13) consolidate into something no single agent authored or maintains, that is, culture as collectively sustained pool of problem-solving operations, inherited by individuals, rescuing each agent from having to derive cognitive protocols *ab ovo*. Culture thus has assembly memory (Step 4), being accumulated constraint-cascade solutions that, once stabilised, become reusable steps channeling subsequent construction without re-derivation.

The progression from situated barter (where each exchange is a full-spectrum negotiation dependent on individual characteristics and relational history) through standardised value norms, through coinage, through markets, through global financial systems... traces such an assembly pathway. The concept of "objective fixed value" is itself an assembly bottleneck, a constraint configuration that coinage, markets, and derivatives all reuse. A market is therefore an assembly with its own integrity requirements — a stable post-singularity phase of the constraint family that governs interactions among agents, transactions, and informational channels (section 1.3.4) — and a collective cognitive light cone (Step 11) encompassing variables — price equilibria, liquidity, inflation... — that no individual participant's cone can reach. These variables are not properties of any individual transaction but of the ensemble whose constraint family admits price-equilibrating dynamics, and whose disruption (a market crash) constitutes a system-level failure irreducible to any single agent's actions.

Culture, as an object with assembly index, simultaneously constrains and enables as any assembly pathway configuration. It prunes individual state-spaces toward configurations compatible with the collective configuration sustaining its persistence — reshaping even attentional and cognitive styles through exposure — while creating conditional dependencies between individuals sharing the same scaffolding. When knowing one individual's state reduces uncertainty about another's — by both being motivated by the same God, governed by the same currency, or bound by the same taboo — the total correlation among individuals

risers above zero (Step 2), consolidating a cultural Markov blanket whose cognitive glue (Step 10) operates through shared meaning (rather than bioelectricity or magnetism, as in other scales, albeit in assembly dependent from these constraint as well). It is within cultural assembly bottlenecks that indexes required for governance assembly become crossable (Sections 3.2.1, 3.2.4).

Step 14 — Bottleneck:

Abstract reasoning, the cultural assembly index towards functional roles

Among all possible cultural assembly configurations, only those passing through the *abstraction bottleneck* lead to paths compatible with governance systems.

What this bottleneck specifically demands is cultural scaffolding of sufficient assembly index to enable context-independent symbolic manipulation, the capacity to operate on categories rather than idiosyncratic particulars, to reason about relations between abstract classes without requiring concrete instantiation. This capacity is not innate but culturally potentiated, as cross-cultural cognitive studies illustrate. When presented with the syllogism "*All Kpelle men are rice farmers; Mr. Smith is not a rice farmer; is he a Kpelle man?*", the Kpelle farmer replies along the lines of "*I don't know the man in person. I have not laid eyes on the man himself.*" The logical deduction, that Mr. Smith cannot be a Kpelle man, is trivial for anyone whose cultural scaffolding supports decontextualised categorical reasoning, yet the Kpelle farmer doesn't run abstract category membership, but rather rely on direct experiential grounding. This is not a cognitive deficit!, but an expression of situated reasoning being optimal for a cultural niche where generalised abstractions provide less adaptive value than contextual, instance-bound knowledge, representing thus a local optimum in cultural assembly space, highly viable within its own problem space, however not on an assembly pathway conducive to higher-order indexes compatible with governance cognitive systems.

A schooled twelve-year-old who resolves the syllogism instantly reflects not superior intelligence, but participation in a cultural assembly that has, over millennia, built and maintained scaffolding infrastructure for context-independent symbolic manipulation. This abstraction reasoning being culturally accumulated assembly memory (Step 3) that potentiates a cognitive capacity optimal for new sets of problem spaces.

What abstraction as a bottleneck demands, is a cultural scaffolding of sufficient assembly index to decouple reasoning from situated experience, to enable operating on categories, roles, and rules as abstract objects independent of the particular individual instantiating them. What cannot exist without crossing this bottleneck are standardised functional roles, such as "judge," "citizen," or "president", crystallised as abstract positions defined by systemic functions rather than individual, idiosyncratic persons and their characteristics. This is a coarse-graining (Step 3) at the human-collective level: many distinct individuals occupy the same macrostate (e.g., "judge") because the partition that defines the role groups together all individuals indistinguishable at the role level — the role's powers, constraints, and relational structure remain invariant across individual occupants. Below this threshold, role, function, selection and maintenance depends entirely on personalised relations and situated reputation. Beyond it, roles persist across individual occupants as properties of the collective assembly itself, as sieve tube cells' transport function (Step 9) persists across cellular turnover. The role of "judge" cannot be understood as a property of any individual who occupies it, it is a compressed node (Step 12) within the collective's bureaucratic architecture, specifying a functional position that channels system-level

constraint propagation regardless of which particular agent fills it. Without this abstraction, the constitutional code of Step 17 would have no addressable architecture to govern, no roles to assign powers to, no offices to constrain, no functional positions to persist across the generational turnover that governance must survive (Sections 3.2.2, 3.2.3).

Step 15 — Bottleneck:

Language as constitutive code

The functional roles crystallised at Step 15 require a substrate capable of encoding, storing, and transmitting abstract specifications across agents and generations, a requirement that no individual agent can satisfy alone. Written language is that substrate, constituting a further critical assembly bottleneck in cultural index. Irreducible to any single agent, language enables memory augmentation through shared external storage, environmental simplification through common labelling that collectively prunes divergent perspectives, coordination through explicit planning that stabilises intentions across agents, transcendence of path-dependent individual learning through idea transfer between minds, and complex data manipulation through written text enabling iterative problem-solving of otherwise impossible sophistication. Each of these functions instantiates the bowtie compression-decompression architecture of Step 12, high-dimensional experiential states compress into symbolic representations (words, sentences, stories, legal texts...) that are stored, transmitted, and decompressed contextually by receivers exercising their own generative models. No agent created language, nor does any agent maintain it, nor could any agent perform the functions it enables in isolation without it. Language persists as collectively maintained assembly memory (Step 4), and like all assembly bottlenecks, once crossed it becomes reusable memory upon which everything downstream depends.

What distinguishes this bottleneck from prior scaffolding, however, is not merely cognitive enhancement but a constitutive property, in that written language is the code in which governance protocols are written. Laws, bills, regulatory frameworks, and constitutional texts exist as linguistic assemblies, that is, formalised constraint architectures whose persistence, transmissibility, and iterability exceed anything oral culture or tacit norm can sustain. This marks the transition from cultural Markov blankets coordinating through shared meaning (Step 14) toward governance Markov blankets coordinating through codified constraint structures — a singularity-node (Step 5) in the cultural-linguistic constraint family at which the joint distribution consolidates into a stable post-transition phase whose collective cognitive light cone (Step 11) encompasses problem spaces, variables, and timescales no individual human's cone could reach. Without crossing this bottleneck, abstract roles (Step 15) cannot be encoded into persistent institutional architectures, coordination protocols cannot outlive the social memory of their practitioners, and the self-referential code structures that define constitutions cannot be assembled (Section 3.2.6).

Step 16 — Bottleneck:

The constitutional assembly, a self-referential code

DNA does not encode protocols for altering its own code, rather changing by mutation, unplanned perturbation subject to external selection. Constitutions, by contrast, encode the primary constraints on agent state-space, but also the (secondary) procedures

for generating future constraints, for modifying both that which is constrained and the constraining procedure itself, plus the criteria to determine when modifications are warranted. This recursive assembly memory — constraint configurations that include instructions for their own updating — enable governance assemblies to maintain identity through transformation, adapting to novel problem spaces without dissolving the Markov blanket integrity (Step 8).

A constitution's ontology is constitutive rather than representational, concerned not with truth but with viability toward higher-level system maintenance. Asking whether one constitution represents reality more accurately than another is as incoherent as asking whether zebra DNA is truer than lion DNA. Both are evaluated against persistence, not correspondence. Error-correction mechanisms, such as judicial review, amendment procedures, interpretive precedent... maintain normative coherence across application cycles, as DNA proofreading enzymes maintain sequence fidelity, correcting drift that would otherwise decouple the system's generative model (Step 6) from its operational environment. Constitutional self-preservation clauses (e.g. the German eternity clause, the Portuguese Article 288) encode integrity preferences at governance scale, the same non-uniform priors over states compatible with persistence that first appeared at the non-equilibrium threshold of Step 5.

Step 17 — Framing:

Humans as embedded agents within the governance Markov blanket

Humans are simultaneously part of the assembly index through which governance was constructed (every bottleneck from cultural scaffolding [Step 11] through constitutional code [Step 15] was assembled through human cognitive operations) and the operational substrate through which governance cognition is instantiated. The ribosome is not external to the gene it expresses, the human is not external to the constitution they enact.

Humans occupy functionally weighted roles (Step 15) within a constitutional protocol (Step 17), bringing their own generative models and integrity drives to positions whose system-level function exceeds their individual cognitive light cones (Step 11), operating through the bidirectional hard protocol / dynamic drive architecture (Step 12). Downward, governance hard protocols condition the lower-level human distribution exactly as Step 6 described: the constitutional state Y projects downward as $P(X|Y)$ on the human distribution, reweighting the alphabet of attainable human microstates — what individuals can attend to, value, pursue... — generating Red Queen dynamics from which agents cannot exit without severe fitness penalties, creating escalating adaptation demands toward system-level compatibility. Upwards, constitutional code does not execute itself, i.e., which possibilities become phenotypically expressed depends on the dynamic drives of humans instantiating them, that is, their interpretive traditions, selective enforcement, cultural resistance... Over evolutionary time, the conjunction of downward conditioning and Red Queen pressures creates conditions for Baldwinian assimilation, with traits aligned with governance-defined roles being systematically favoured by the path-reweighting selection mechanic of Step 5 — applied here to evolutionary paths of the human substrate — while capacities outsourced to the collective atrophy, and constitutional code cascades towards DNA code, with governance systems in the loop engineering their own substrate. The constraint-enablement signature (Step 10) holds with integration constraining individual possibility while enabling collective access to assembly pathways no individual could reach. The governance system shapes the humans who interpret it, while those interpretations reshape governance

expression, and in turn that expression reshapes selective pressures on humans themselves. The recursion has no outside within the Governance Cognitive System markov blanket (Sections 4.3).

Step 18 — Final Framing:

The ontological status of Governance Cognitive Systems

Governance systems satisfy the same informational criteria that define cognition wherever it appears in the assembly continuum. Constraint propagation compressing possibility spaces, generative models encoding predictions against which observations are evaluated, active inference minimising prediction error through perception-action cycles, and integrity drives maintaining preferred states against entropic decay. Though the substrate changes from what is usually the object of cognitive science — to documents and legal architectures, rather than bioelectric networks — information as constraint propagation is substrate-independent, and the cognitive operations enabled are defined by their active constraint towards integrity functional role.

Today *Governance* studies are sitting in university departments as if categorically different from the systems studied by the natural sciences. The principles that consolidate a tissue-level Markov blanket (Step 8) and those that consolidate a governance-level Markov blanket (Step 15) are the same class of assembly bottleneck, both produce higher-level cognitive systems whose properties belong to no constituent, with the difference being one of assembly index.

Given that governance systems are cognitive systems, they possess analysable cognitive properties. Variable sensitivities determined by rendered metrics, blind spots from compression artifacts (Step 10), bounded cognitive light cones (Step 9), biases existing only at their organisational level. A governance system lacking variable sensitivity to biodiversity is cognitively blind to it, as structurally as a visual system without ultraviolet photoreceptors. Recognising this opens governance to the same vocabulary of cognitive diagnosis and architectural design that the biological sciences bring to the systems they study. Cognitive Science and Political Science, then, do not study different phenomena requiring different epistemologies, but rather study the same phenomenon at different assembly indices, substrates, across different problem spaces, yet by the same principles. And thus should seek to explicitly inform each other.

Part III

Filling the Gap

Chapter 5 — What the state of the art is missing

5.1 Limitations in Cognitive Science, Biology and Assembly Theory

5.1.1 A Pioneering, Yet (self)Constrained, Foundation

Minsky's *Society of Mind* (1986) constitutes the historical seed of the agentic, scale-free conception of cognition that Section 2.3 articulated as foundational. Yet, examining it from the vantage point of the framework now established, it reveals both an anticipatory richness and a self-imposed structural limitation.

First and foremost, it should be stated that the author explicitly prescribed against reading beyond the *Society* metaphor³⁴³, regardless of having laid out the logical pillars that allow for it to be extended. However, when one reads *Society of Mind* — while kindly disobeying the author's self-imposed limitations — it conveys inspirations for conceiving of Governance as a *Cognitive System*.

That is, Minsky's focus on relationships between *agents* foreshadows the *externalistic* and *distributed* paradigm that would later gain prominence. This intuition isn't the only future oriented insights offered by the author, for instance, the book acknowledges the role of emotion in cognition, stating that "*our culture wrongly teaches us that thoughts and feelings lie in almost separate worlds*", while also hinting at a broader ecological cognition, with observations such as "*no one really learns alone, since every step employs many things we've learned before, from language, family, and friends*". The *Society of Mind* thus contains a breadth of view that often transcend the cognitivist paradigm Minsky primarily occupied — reflecting a deep understanding of the complex, interwoven nature of cognitive processes — however simultaneously being largely limited by the concerns and approach of its time: being overly rigid and relying on hard-coded rules and representations³⁴⁴, while being focused solely on internal processes within individual skulls.

These limitations become all the more apparent when Minsky makes various efforts to compare the mind to a society³⁴⁵ — where agents' organisation and labour division constitute mental processes — while explicitly cautions against drawing too close an analogy between the low-level agents of a single mind and the members of a human community⁷⁷, stating that "*Because our little mental agents are so limited, we should not try to extend very far the analogy between them and human supervisors and workers*".

This caution is needlessly restrictive especially when Minsky inadvertently lays a lot of groundwork for understanding societal cognition, as his book is replete with ideas that implicitly theorise its foundational pillars: indirectly supporting the externalist, distributed, scale-free and substrate independent nature of cognition, as developed across Sections 2.3 and 2.7.

The author makes it clear that what matters is the functional equivalence and the preservation of processes, not the material composition, bypassing the need for biology and operationalising cognition³⁴⁶ to concrete sets of functional processes. All the while introducing a scale-free, distributed, agentic notion of cognition. Thus, Minsky unwittingly — when taking his premises to their logical extrapolations — significantly contributes to *Societal Cognition* as the sort being explored in this thesis. However, the author doesn't seem to take the logical order of his ideas into the full depth of their *societal cognition* implications. To conceive of *agents* so as to encompass societal cognition, one ought to take on Minsky's ideas to their next logically derived step, in a manner fully supported by the author's assumptions, by extending these agents to beyond the borders of the skull, running and

³⁴³ "we should not try to find a close analogy between the low-level agents of a single mind and the members of a human community" (Minsky, 1988)

³⁴⁴ "GRASPING agents want to keep hold of the cup. BALANCING agents want to keep the tea from spilling out. THIRST agents want you to drink the tea. MOVING agents want to get the cup to your lips." (Minsky, 1986)

³⁴⁵ "minds are like societies" (Minsky, 1986)

³⁴⁶ "as far as I'm concerned, the so-called problem of body and mind does not hold any mystery: Minds are simply what brains do. Whenever we speak about a mind, we're speaking of the processes that carry our brains from state to state." (Minsky, 1986)

articulated across a variety of societal embodiments³⁴⁷. A step that the author, in various passages, seems comfortable in taking, either stretching his ideas to a societal scale, or making cybernetic analogies at the societal level:

“The human brain contains so many agencies and connections that it resembles a great nation of cities and towns, linked by vast networks of roads and highways.”

(...)

“[Negative Feedback] Supply a “summary” device that estimates the total activity in the agency and then broadcasts to that agency an “inhibitory” signal whose strength is in proportion to that total. This will tend to damp down incipient avalanches. (...) All human communities seem to work out policies for how their members ought to think, in forms that are thought of as common sense or as law, religion, or philosophy.”

(Minsky, 1986)

When taken at the level of Social Cognition, some of Minsky’s ideas scale so well so as to be insightfully novel. For example — when read with this renewed horizon of interpretation — the idea that *“Societies make laws adapting and changing themselves”* illustrates the adaptive, *neuroplastic* nature of societies, similar to cognitive systems that learn and evolve by changing the contents of their own associations³⁴⁸. In fact, the notion of *social protocols* as dynamic and self-regulating cognitive systems can be read as a logical extension of the author’s premises across the *Society of Mind*, and inspiration from Minsky’s symbolic ideas will populate further sections of this thesis.

Beyond its *ethos* constraint, further limitations can be observed on the materiality of ideas both present and absent from Minsky’s book, revealing deeper structural limitations regarding the materiality and dynamics of the agents Minsky theorises. As an approach, the overall stance of the book is characterised by Minsky’s suspicion of subsymbolic solutions, relying on hard-coded rules and representations to the detriment of a more statistical, network based, bottom-up, state-space and free energy minimisation approach. This isn’t surprising given Minsky’s AI position³⁴⁹. Seymour Papert — Minsky’s co-author on *Perceptrons* (a book criticising aspects of connectionism and the subsymbolic approach, specially highlighting limitations of the single layer architecture) — even made the rhetoric question *“Did Minsky and I try to kill connectionism, and how do we feel now about its resurrection?”*³⁵⁰ (Minsky & Papert, 2017), revealing an awareness of both the hard avoidance of non-representational solutions³⁵¹, and its counterproductive stance. This attitude is reflected on the lack of dynamism offered by *Society of Mind*’s solutions: its cognitivist notion of the mind as a functional decomposition of narrowly defined tasks defies

³⁴⁷ Plus, importantly, granting *agents* within inherent cognition, nested within higher level cognitive Markov Blankets — see chapter 2, A Kaleidoscope of Selves subsection (Section 2.4)

³⁴⁸ Resonates with other within-the-skulls assertions by the author, such as *“‘Learning’ is making useful changes in the workings of our minds.”* (Minsky, 1986)

³⁴⁹ Minsky did point out, in a rigorous and structural manner, fundamental flaws of early architectures — such as single-layer networks handling of non-linearly separable problems, like XOR (Minsky & Papert, 2017) — which nudged the field to evolve and thus promoting future solutions);

³⁵⁰ To which he replied: *“Something more complex than a plea is needed. ...[P]art of our drive came ... from the fact that funding and research energy were being dissipated on what still appears to me ... to be misleading attempts to use connectionist methods in practical applications”*

³⁵¹ Although, when designing solutions entailed by the notion of governance as a cognitive system, for these to be transparent and democratic, they ought to have a strong symbolic leaning (so as to be understandable and broadly represented by the constituents it serves)

more sophisticated integrated approaches such as enactivism, or active inference, in which a dynamical and plastic relation with the environment trumps hard coded rules, lacking the autonomy to dynamically interpret bowtie-bottlenecks in context specific ways, as elaborated in Section 2.5. Even within the scope encompassed outside of more dynamic paradigms, the book lacks models or concerns for such fundamental entities as *Attention*.

Regardless of the *Society of Mind* being informative and constructive to the conception of a broader cognitive system — as aimed by this thesis — various limitations impede it from exhausting the breadth and depth of concepts and tools required to uphold *Governance as a cognitive system* on its own.

Furthermore, although the author's conception of *agents* seems aligned with *basal intelligence's* principles, so as to “*understand how Mind scales—how large numbers of competent subunits can work together to become intelligences that expand the scale of their possible goal*” (Levin, 2023), this done in a manner which not only neglects, but outright rejects the inherent intelligence of its subunits. This goes against current yieldings in the study of *basal intelligence*, which — by recognising the intrinsic intelligence of the *mind's subunits* — hints at emergence through the collective actions of autonomously intelligent agents (Fields, Glazebrook, & Levin, 2021). Breaking through a simplistic and mechanistic notion of agent is fundamental to envision individual human nodes as intelligent components within a governance cognitive system (as developed across Sections 3.2.4, 4.2 and 4.3), a notion which isn't afforded in Minsky's narrow definition of *agents*.

This rejection of subunit intelligence is not merely a definitional choice but has substantive consequences for the kinds of entities that can occupy the agent role. Take the example of *Physarum*, a unicellular slime mould, that when positioned in an arena with various inert glass objects of differing mass distributions, employs a biomechanical mechanism to process environmental information and consistently grows toward the heavier object (Murugan et al., 2021). Such an entity, exhibiting volition, valence, and adaptive problem-solving without neural architecture, sits squarely outside Minsky's available conceptual repertoire, in which an agent is reduced to a mere executor of scripted functions.

As per left off above, Minsky's agents are stripped of volition, valence (the preference or avoidance of particular stimuli or conditions) and cognition, performing narrow tasks from which *true cognition* emerges at the *bureaucracy* scale over a protocol — thus lacking inherent agent level *preferences*, motivation and active homeostatic strategies — being mere executors of scripted functions³⁵². This symbolic executor, *hard-coded* static view overlooks the real advantages of having dynamic agents with subunit autonomy while also severely limits the kinds of entities that can perform the role of *agent* in medium, type and scale³⁵³.

This subsection will be informed primarily by biological systems, so by subsystems here one can understand organs, tissues, cells, or even smaller sub-agencies within the cell (as we'll see). However, I suspect that true groundbreaking advances in Artificial Intelligence, Engineering and Medicine wait at the cusp of yielding the advantages provided by nested cognitive, valence agencies. It is my belief that agential design and agentic programming — instead of hard-coded strict rules based engineering, static fixed materials, or intrusive mechanic level pathways — in which the material and its derived properties are agentic, is the path towards which much of the human scientific and engineering progress will catch up with biological solutions already following this formula, towards great advancements.

³⁵² As per Section 2.3 (From collaborator to generator) in Chapter 2

³⁵³ Given that *Minsky's agents* individually lack *true cognition*, would one be inclined to regard an individual human as a societal *agent* on a governance bureaucracy?

The cumulative weight of these limitations points to a single, deeper omission underlying them all. Minsky's main oversight is not addressing the *why* of Cognition at a fundamental level and having not formalised it informationally with regards to a system's assembly and maintenance. Specifically, Minsky offers no probability-space foundation for what an "agent's state" formally is (compare Section 1.1.1), no criterion analogous to positive total correlation (Section 1.1.3) for diagnosing when a collection of agents constitutes a system versus a mere aggregation and no formal apparatus for level transitions of the kind developed at singularity-nodes (Section 1.3.3). Without these, Minsky's "emergence" remain at the intuitive or at least non first-principles level.

Minsky's original bureaucracy is correct in highlighting how many small processes, each narrower than the whole, scale to emergent higher level cognitive processes, but it misses the potent synergy of subunit adaptiveness and local intelligence that *Basal Cognition* spotlights, as well as the criteria by which selves are delimited (addressed in Section 2.4 through Markov-blanket statistical individuation and grounded in the recursive coarse-graining apparatus of Section 1.1.4) and how they are informationally defined. So each agent should be conceived as not passively obeying a strict script, but in experimenting, learning and holding transient generative models, to the degree these terms apply to their level, all to strive for local (agentic) and general (system wide) homeostasis (as developed in Sections 2.1 and 2.2). It also humbles the anthropomorphised view of intelligence as only present at bureaucratic scales, akin to those we occupy, extending it to smaller levels nevertheless equipped with internal states, plastic responses, and memory³⁵⁴.

5.1.2 Scale-Free Yet Not In Assembly's first principles

Where Minsky's framework falters at the level of agent dynamism and informational anchoring, Michael Levin's basal cognition programme advances on these dimensions, recognising the inherent intelligence of subunits and extending cognition across biological scales. Levin's contributions, as Sections 2.2 and 2.4 attest, are foundational to the framework developed in this thesis. Yet, examining basal cognition from the vantage point of the assembly–informational synthesis articulated in Chapters 1 and 2 reveals a complementary set of limitations. Although being less biased towards higher order biology, basal cognition still isn't articulated with Assembly Theory nor with the probability-space, total-correlation, and singularity-node apparatus of Chapter 1 to the point of characterising cognition informationally as developed in this thesis.

While the section within this chapter is aimed at laying out the route cognitive science has taken to incorporate the environment while *externalising* the self, reframed in this thesis to also provide a perspective on potential higher level *cognitive glues*³⁵⁵ linking lower level human agents into higher-level cognitive functions, I would like to first remark that — as per the methodology committed to in Section 2.7 — the *Basal Cognition* approach does not exhaust my thesis of regarding Governance systems as Cognitive Systems. While stating that cognition is takes place across scales does by principle encompass agglomerates at the scale of a society to contain cognitive functions, it does not, however, regard the protocols of governance as cognitive architectures; it does not put emphasis on the concrete cognitive processes, affinities and compatibilities afforded per types of protocol at the level of

³⁵⁴ "[in Cellular signaling] protein pathways and transcriptional networks exhibit interesting and unexpected capabilities such as trainability (memory) and information processing in a context-sensitive manner. Specifically, they may be amenable to manipulation via their history of stimuli (equivalent to experiences in behavioral science)" (Mathews et al., 2023, p.1)

³⁵⁵ Further defined and illustrated on the next section

governance assemblies (neither does it conceptualise assemblies nor assembly index, thus not conceiving of institutions as such, nor their cognitive role in higher level societal cognitive assemblies³⁵⁶); likewise it does not contemplate defining the *protocol* and *bureaucracy* as a processing code from which designed cognitive processes take place; or institutions as cognitive modules, and constitutions and such foundational societal protocols as the bottle-neck code for higher-level cognitive process enabled by Governance Cognitive Systems. All these dimensions have been developed across Chapters 3 and 4.

The implication is significant. Basal cognition, while indispensable for grounding the scale-free cognitive thesis advanced in Chapter 2, stops short of the informational and assembly-theoretic apparatus required to extend its insights into the governance domain. The biological substrate it primarily theorises — bioelectric networks, cellular collectives, regenerative morphogenesis... — provides an existence proof of multi-scale cognition without supplying the reductionist substrate developed in Chapter 1³⁵⁷. This thesis treats Levin's framework not as something to be displaced, but as an empirical and conceptual foundation that, when articulated with Assembly Theory and the probability-space apparatus and path-reweighting selection mechanic of Chapter 1 (particularly Sections 1.1.1, 1.1.3, 1.3.3 and 1.3.4), becomes capable of bridging cellular and governance scales within a single cognitive vocabulary.

5.1.3 Analytical, yet Without Ontological Commitment

If Minsky's limitation is internalist rigidity and Levin's is the absence of informational formalisation, Edwin Hutchins's distributed cognition addresses the externalist front by widening the unit of analysis to encompass socio-technical systems. As Section 3.1 elaborated, Hutchins's framework provides crucial scaffolding for the analytical move from individual to system-level cognition. However, three complementary limitations prevent it from reaching the framework articulated in this thesis: an analytical-only stance that withholds ontological commitment; an unrealised fractal insight; and a holistic dissolution of system boundaries.

First, as an epistemological prologue, I would like to state that Hutchins regards Distributed Cognition as an analytical framework without aspiring to an ontological claim³⁵⁸. As per Hutchins (2013), "*There is no series of experiments or set of observations that could prove or disprove the distributed cognition perspective, because the perspective itself makes no empirical claims*". Whereas this thesis regards the reality of *Cognitive Governance Systems* as an ontological one, capable of sustaining numerous empirical hypotheses that can be confirmed or rejected through investigation, with the criteria for systemhood³⁵⁹ for level consolidation³⁶⁰ and for selection³⁶¹ all being checkable in principle, against the

³⁵⁶ As showcased on the third synthesis chapter

³⁵⁷ Information as the measurable departure of a joint distribution from its maximum-entropy reference on whatever probability space the level of inquiry requires (Section 1.1), the three complementary registers through which this is registered (negentropy, mutual information, and total correlation, Section 1.1.2); the recursive coarse-graining grammar that grounds the hierarchical and nested picture of levels (Section 1.1.4), and the assembly memory dynamics through which constraint configurations accumulate as historically realised bottlenecks (Section 1.2)

³⁵⁸ "*It is important to underscore that Hutchins' position is that distributed cognition is an analytical framework and not an ontological claim*" (Nersessian, 2019, p. 556).

³⁵⁹ Positive total correlation, Section 1.1.3

³⁶⁰ Singularity-nodes as non-analyticities of constraint families or bifurcations of the information-bottleneck Lagrangian, Section 1.3.3

³⁶¹ Path-reweighting of probability mass over assembly paths, Section 1.3.4

structure of the relevant joint distributions, as the later section on *Future Research* will attest to.

A second limitation appears at the level of fractal connectivity, where Hutchins glimpses but does not pursue the level-traversing implication of his own insight. As it should be acknowledged that Hutchins did integrate the general conjecture that cognition follows a fractal structure of connectivity. He posited that "*cognitive systems at all scales exhibit a fractal structure where dense nodes of interconnectivity are separated by sparser connections*" asserting that, "*I do not expect intelligent behavior to arise from networks having uniform connectivity*" (Hutchins, 2013). However, while this insight could logically entail the possibility that structures that are external at one level of analysis become internal at a higher level — a possibility the recursive coarse-graining apparatus of Section 1.1.4 formalises explicitly³⁶² — this line of reasoning is never explicitly made so as to regard the protocols, embodiment, active-inference and maintenance of Governance Systems as Cognitive

The third limitation concerns the absence of criteria for delimiting one cognitive level from another, which leads Distributed Cognition to dissolve, instead of identifying discrete higher-order assemblies. Given that *Distributed Cognition* did not provide any criteria to delimitate levels, or orders, of *fractal-like* cognitive systems, instead it dissolved delimitations by considering overall ecological assemblies as cognitive — defaulting to *Cognitive Ecology*³⁶³ — composed by individuals, groups and their relation, cultural practices, artefacts, and an environment that form mutually supportive networks³⁶⁴ that extend so as to entail the overall environment³⁶⁵, therefore adopting an holistic perspective on cognition³⁶⁶, thus foregoing an assembly – based on actively maintaining integrity and homeostasis – perspective on cognitive processes. Whereas this thesis offers concrete³⁶⁷ (and I would argue *useful*) metrics and concepts whereby to attest the cognitive light cone of a given system in the scope of variables and spatiotemporal horizons of its actions, feedback error and world model. Thereby delimiting it cognitively in relation to the goals promoting its level integrity, while also, physically hinting at its embodiment horizons — to the extent of the Cognitive glues binding agents together, the effectiveness and extension of its code applications, and primarily on the maintenance of "singularity node status" (Section 1.1.3) — compared to uncoupled individual human agent degrees of freedom.

"As the field of cognitive ecology grows and matures, we will achieve a better understanding of why cognitive processes are as they are. This will be true both in a synchronic sense in terms of the functional relationships among the elements of the contemporary cognitive ecology, and also in a diachronic sense where

³⁶² Where each higher level is one function-composition further from the base than the level below and where the same underlying probability space supports a nested hierarchy of descriptions

³⁶³ "Cognitive ecology is the study of cognitive phenomena in context. In particular, it points to the web of mutual dependence among the elements of a cognitive ecosystem." (Hutchins, 2010, p. 705)

³⁶⁴ "Cultural practices are emergent structural configurations in a rich network of relationships" (Hutchins, 2013)

³⁶⁵ "An understanding of cognitive phenomena must include a consideration of the environments in which cognitive processes develop and operate." (Hutchins, 2010, p. 706)

³⁶⁶ "As the field of cognitive ecology grows and matures, we will achieve a better understanding of why cognitive processes are as they are. This will be true both in a synchronic sense in terms of the functional relationships among the elements of the contemporary cognitive ecology, and also in a diachronic sense where understanding the evolution of cognition will be recast as understanding the evolution of cognitive ecologies." (Hutchins, 2010, p. 712)

³⁶⁷ Through the singularity-node criteria of Section 1.3.3, the Markov-blanket criteria of Section 2.4, the functionalist methodology of Section 2.7, and the considerations on Cognitive Glue developed in Sections 2.7 and 3.2.4

understanding the evolution of cognition will be recast as understanding the evolution of cognitive ecologies.”

(Hutchins, 2010, p. 712)

The convergence of these three limitations leaves Distributed Cognition unable to identify the assembly bottlenecks and Markov-blanket statistical signatures — both of which depend on the underlying probability-space machinery of Section 1.1 — by which higher-order cognitive systems can be ontologically delimited. This is not a deficiency of Hutchins's analytical project on its own terms, but rather a structural ceiling that the present framework is designed to lift.

5.1.4 The Persistent Anchoring to Individual Cognition

The remaining branch of the externalists — Extended and Scaffolding Cognition (Clark & Chalmers, 1998; Clark, 1997) — sit between Hutchins's wide-scope distribution and traditional internalism, as Section 3.1 examined. Its parity-based approach productively bridges the inside/outside divide, but, like its predecessors, retains a centre of gravity that prevents it from reaching the higher-order assemblies this thesis articulates. Three limitations follow in sequence: the parity principle's anchoring to the individual orbit, the resulting unresolved scoping problem, and the perpetuation of these limitations once Active Inference is incorporated.

Intentionally, my approach aims to solve a lacuna not explicitly resolved by *Extended* nor *Scaffolding* Cognition. Given that both these frameworks assume a central cognitive unit³⁶⁸ (an assumption polar to the *overgeneralization* of *Ecological Cognition*), they end up restricting the *Parity Principle* to the orbit of the individual's capabilities. Yet the framing of *individual* remains at human-level. Thus, a concept like *Dissipation of Reason*³⁶⁹ — in which *reason* would be regarded by *Ecological Cognition* as an overall property *emergent* from the various interactions in the environment — is read by *Extended Cognition* as the myriad of support systems to which a person offloads cognition, with *reason* being found in the paired coupling between the individual and its scaffoldings, which limits the kinds of cognitive operations and cognitive models to those available at the sphere of the individual. This thesis approach, however, allows go beyond the *parity principle* anchored to individual Human cognitive center model, into novel cognitive capabilities, modalities and representations unavailable to the *individual human agent* and only present within the societal, governance scale³⁷⁰, (as articulated in Sections 4.2 and 4.3). Adopting this perspective makes it so that it becomes inaccurate to regard societal achievements — such as the moon landing — as enabled by the scaffoldings supporting an individual aspirational astronaut's desired goal of going to the moon.

Such societal goals result from a higher level operation irreducible to any single individual, and are processed and coordinated at a scale outside of the scope of any

³⁶⁸ “extended mind assumes a center in the cognitive system: the organism (or the organism's brain), which is the normal mind container with respect to which cognition can be said to extend” (Hutchins, 2013, p. 3)

³⁶⁹ “Advanced cognition depends crucially on our abilities to dissipate reasoning: to diffuse achieved knowledge and practical wisdom through complex social structures” (Clark, 1997, p. 180).

³⁷⁰ It is arguable whether, by studying individual neurons, their relationship with other neurons and their environment, yet without acknowledging the higher level brain as a system, notions fundamental to psychology, such as trauma or qualia, would appear.

individual human, with cognitive modalities treating variables only rendered at the inter-human level³⁷¹.

This anchoring to the individual has a structural consequence. The field is left with an unresolved scoping problem that bridges directly into the question of disciplinary integration. Without the framework provided within this thesis, this divergence between External and Ecological — within the spectrum of "individual-plus" to "social-only" — endures as a scoping problem. The resolution offered here treats this not as a binary ontological dilemma but as a level-choice problem within the recursive coarse-graining apparatus of Section 1.1.4, in which individual cognition and distributed cognition are different coarse-grainings of the same underlying probability space, each level legitimate, with the higher-level probability content following by pushforward from the level below. Addressing this scoping problem is the best way to frame — within the realm of Cognitive Science — the need for the inclusion of Political Science's concerns and concepts into *Cognitive Science's* vocabulary, once the higher level assembly of *Governance* is understood from an assembly and cognitive light cone perspective, within a shared epistemology. From the Political Science side, the articulation with cognitive disciplines is simultaneously the best way to both leverage insights into better protocols of decision-making from the areas studying it, and of better grasping the first-principles and mechanics underpinning the nesting of individual to higher level goals and its implications.

Consequently, most of the usefulness in disregarding the nature or location of cognitive modalities in relation to their system — a hallmark of *externalism* — erodes once one steps outside the traditional, anthropomorphic reference point of analysis. As per the language example given above (let alone the Minsky, Levin updated view of *Agent*), there are plenty of instances of structures that only seem distributed at a given level of analysis, becoming embodied constituents of the cognitive system once a higher level view is applied. Given that, even within the traditional human level of analysis, cognitive functions remain distributed — for example, between modalities (in the same way one could state that my intracranial cognitive function is distributed between brain modules) — the usefulness of denoting cognitive processes as taking place *outside*³⁷² is mostly useful towards countering a biocentric bias which needed not even be addressed if one adopted a basal cognition frame of reference to begin with, in which non-biological elements (such as phosphate³⁷³ or calcium³⁷⁴) are more immediately and clearly related to cognitive processes.

Also, outside of what? The body? If one implants a chip, or has its genome edited, are the modalities played by their function endogenous or exogenous? If I select my progeny's genome by means of designer baby tech, is their unpaired, uncoupled reason endogenous or external, given that a team of biologists and engineers worked towards enhancing their reasoning capabilities?

³⁷¹ . It might even be that no individual human node shares the societal goal being pursued by the Governance Cognitive System (e.g. a war to which no individual recruit wants to be conscripted). In the precise vocabulary of Chapter 1, the joint distribution over the human agents involved sits at a stable post-singularity phase of its constraint family (Section 1.3.3), whose macrostate alphabet — the system-level outcomes those agents jointly produce — is not the Cartesian product of the agent alphabets and cannot be read off any individual's marginal distribution, exactly the bit-parity-like signature of system-level structure invoked in Section 3.1.

³⁷²

³⁷³ "Phosphorylation and dephosphorylation, which occur at different rates at different sites, may affect transcription directly or as one step in a signaling sequence"; "The high-energy state created by modification of the RR [response regulator, to which the phosphate group is transferred] constitutes a marker of environmental conditions at a particular time, which lasts until the chemical group is removed, thus constituting the basis of molecular memory" (Lyon, 2015, p. 5) - "[]" added by me

³⁷⁴ "Calcium/calmodulin-dependent kinase II (CaMKII) [an enzyme activated by calcium ions (Ca²⁺)] is a serine/threonine kinase critical for memory formation in humans. (...) Does it have a prokaryotic homologue? To my knowledge, no one is looking" (Lyon, 2015, p. 2) - "[]" added by me

Beyond the scoping problem itself, a further limitation surfaces when Extended Cognition incorporates Active Inference, where the niche-agent symmetry inherits and amplifies the framework's individual-anchoring. Maintaining such current extended or ecological frameworks perpetuates limitations, such as the problem (as laid out at on my critique on the passiveness of the environment in relation to *Extended Active Inference*'s niche-agent mirror in Section 3.1) of taking an active inference perspective on Ecological Cognition; or considering higher level cognitive modalities irreducible and without correspondence to any given human as loci of cognition; and thus having no framework from which to anchor, define and model generative models sustained at an informationally delineated *environment level*, which nests individual human cognitive light cones along with sets of linguistic, institutional and cultural artifacts as cognitive blue, thus relegating actions, processes and behaviours of higher assemblies entities — such as states, nations and institutions — to the realm of Political Science alone. A unified model is possible once a more beyond-disciplinary encompassing approach, derived from first-principles, converges on their phenomena of interest. When onboarded on this perspective, it becomes clear that once the *scoping problem* is transformed into a *level*³⁷⁵ *problem*, it becomes much less of a problem, along with its *misattributing* problem.

5.2 Limitations in Political Science

5.2.1 The Inheritance

A dualist architecture — as the operative assumption that institutions, constitutions, and the populations they govern are categorically distinct entities, separable for analytical purposes and ultimately differentiable in their ontology — recurs across (at least³⁷⁶) the post-war canon, despite profound disagreements in method and commitment.

Hans Kelsen's pure theory of law sought normative purity by insulating law from sociological fact³⁷⁷. H.L.A. Hart's rule of recognition located the legal system's identity in the internal acceptance of officials, treating that acceptance as a conceptual primitive, far from contemplating it as an embodied cognitive state requiring further analysis³⁷⁸. John Rawls reasoned about justice from an original position whose veil of ignorance abstracts away those embodied, cognitive and cultural conditions through which any actual governance system operates³⁷⁹. Jürgen Habermas developed communicative rationality as a procedural domain whose validity claims float free of the neurocognitive scaffolding that enables them³⁸⁰. Rational choice theory took preferences as exogenously given parameters without of understanding them as configurations actively shaped by the very institutional architectures whose effects they purportedly explain³⁸¹. New institutional economics (North,

³⁷⁵ Of choosing the appropriate coarse-graining (Section 1.1.4)

³⁷⁶ As per what was studied during my bachelor in Law, and for self-containment of this thesis, my analysis start past world war two

³⁷⁷ (Kelsen, 1967)

³⁷⁸ (Brown, 1963)

³⁷⁹ (Rawls, 1971)

³⁸⁰ (Gaspar, 1999)

³⁸¹ The exogenous-preferences assumption is defended by Becker and Stigler (1977) forming the pillar of rational choice theory. Bowles (1998) provides the standard critique (from economics), illustrating that institutions shape the preferences whose effects they are invoked to explain. The framework developed in this thesis takes this critique even further, with preferences not merely being shaped by institutions, but co-constituted with them through cultural scaffolding (Section 3.2.4), downward causation from hard protocols (Section 4.3), and Baldwinian assimilation over evolutionary time (Section 4.3). Treating preferences as

1990) characterises institutions as "the rules of the game" external to the players who navigate them. Although the differences among these frameworks are profound, their commonality is nevertheless structural. Each presupposes that the appropriate level of analysis for governance can be specified independently of the multi-scale informational architecture through which that level actually persists³⁸². This same architectural assumption organises everyday political thought. I will illustrate and deconstruct it using Luhmann's framework, both due to its influence and sophistication, and because it foreshadows several insights this thesis develops while highlighting the major dualistic flaw of post war mainstream political thought.

In a delightful yet orthogonal *parallel* with Shannon³⁸³, Luhmann too leverages communication, viewing it as the operative element of system maintenance. Leading with it towards the recognition that systems must process perturbations through their own internal logic. The critique that follows is therefore not directed against systems thinking in political analysis, which this thesis explicitly embraces, but against the specific architecture through which Luhmann and the broader tradition he exemplifies have instantiated that thinking while perpetuating a cartesian dualism. Especially since this thesis – despite the novelty of its conclusions and the monist first-principles framework of its approach – is far from the first endeavour to leverage Biology and Cognitive Science in the study of governance. Owing in scope to Luhmann's ambitious project of Social Systems Theory (1995), in which he borrowed and adapted the lens of *autopoiesis* from biology, to develop a distinct and prominent philosophy through which he conceptualised the self-production and maintenance of social systems.

Autopoiesis (literally meaning "self-creation" or "self-production") was developed and introduced by Humberto Maturana and Francisco Varela (1980) – employing system's thinking to understand living entities as dynamical networks of molecular processes and transformations that continuously produce the component that realise them, all while interacting with its environment and maintaining its distinct organisational identity³⁸⁴ – thereby describing living systems as organisationally closed³⁸⁵ (despite material flux) and self-producing. Niklas Luhmann significantly broadened this biological concept of autopoiesis so as to apply it in the domain of social systems (1995). In Luhmann's framework, society and its systems are themselves autopoietic, that is, operationally closed and self-referential entities that produce their own elements through their own operations. These characteristics extend to all of society's systems, from law to politics, economy to science³⁸⁶, each operating according to their own binary code (legal/illegal, government/opposition, payment/non-payment, true/false) and reproducing itself through their own set of recognised operations that reference only their own priors. The legal system, for instance, determines what is legal or illegal only by reference to its own prior legal decisions and statutes, in turn

exogenous is thus another instance of the dualist inheritance this section showcases, with inputs and outputs of governance artificially separated instead of recursively co-produced.

³⁸² As per Section 1.1.4, each implicitly fixes a coarse-graining without acknowledging that the coarse-graining itself is a choice whose probabilistic content is determined by the chain of compositions that connects it to the levels beneath. And according to Section 1.3.4, each treats higher-level structure as ontologically autonomous rather than as the conditioning of lower-level distributions by higher-level states

³⁸³ One need not great theory of mind to triangulate my admiration towards and inspiration from these characters

³⁸⁴ As they articulated, "*An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relationships) that produced them; and (ii) constitute it (the machine) as a concrete unity in space in which they (the components) exist by specifying the topological domain of its realization as such a network*" (Maturana & Varela, 1980, p. 78 - 79)

³⁸⁵ In the sense that they themselves produce and reproduce their own components.

³⁸⁶ His vast biography exploring these areas include: *Essays on self-reference* (1990); *Law as a social system* (2004) and "*Die Wirtschaft der Gesellschaft*" (*The Economy of Society*) (1988).

creating said statutes by the proceedings of further laws, which subsequently can serve as prerequisites for subsequent operations – and thus sustaining legality by self-creation as opposed to references in morality, economics or personal character – in a closed, self-sustaining manner.

Although the role of binary codes in maintaining their self-referential operations (i.e. communication, as seen below), recurs throughout Luhmann's writings – such as the binary code of the legal system as legal/illegal is being analysed in *Law as a Social System* (2004) – the notion of code itself, Luhmann (1982, p.168) states *"A code, accordingly, is a structure with specific characteristics: it has a special capacity for guiding selections. It is able to locate a counterpart for each relevant item that falls within its domain. / We can clarify the way a code functions by recalling the special sort of code that originally made the concept fashionable: the one used for sending messages. Above and beyond this technically special case, however, the principle of assigning a counterpart to every item in a given field has a more general significance for the structure of complex systems. The morphology of simple organisms, for example, already requires it. Apart from code failures or mutations, this principle makes possible a strict coordination between mechanisms for genetic reproduction and the elementary structure of organisms. This coordination, in turn, makes organic evolution possible. Now, for meaning-generating systems, language is the primary code."*

As Luhmann famously stated, *"The legal system operates in a normatively closed and, at the same time, cognitively open way"* (2004, p. 106), meaning that, although the legal system can translate *perturbations* from the external environment, it must incorporate them into its own language of legal/illegal to become permeable and rendered in its medium. Generalising this means that social systems can only process information that they themselves generate internally. And it is through this operational closure that systems distinguish themselves from the environment, creating their boundaries not as physical demarcations but through the ongoing differentiation of its specific mode of communication from other forms of communication (or non-communication). Crucially *communication* would not be a mere medium in Luhmann's approach, but the embodiment of the system itself. As Luhmann stated, *"Social systems use communication as their particular mode of autopoietic reproduction. Their elements are communications which are recursively produced and reproduced by a network of communications and which cannot exist outside of such a network"* (1995, p. 174). Thus, Luhmann's theory posits that social systems are not composed of human individuals or their actions, but constituted entirely of communications.

Making said network structures of *communication* the autopoietic element and defining *non-communication* (in the narrow sense) as the system's environment, thus making humans exist in the environment of social systems and not as their components. Governance would not then emerge from human will, or collective decision-making, but from the autonomous operational logic of differentiated functional systems. Furthermore, for Luhmann, the system itself – being constituted by fleeting communicative events – had no persistent actors or memory structures in the stable traditional sense, meaning that the system is not *stored* but is *generated* and carried forward by an ongoing chain of communications (Luhmann, N., 1995).

However operationally closed, systems would nevertheless be subject to being "irritated"³⁸⁷ by each other – akin to external stimuli being integrated via internal pathways, as per the internal system's logic – becoming then available for *structural coupling* (Luhmann,

³⁸⁷ Akin to how the nervous system can integrate stimuli from the outside, by converting into its own operating logic external input via its own sensing logic and mechanisms

N., 1995) thus influencing each other without violating their respective closures and self-referential communication protocols. Reacting through their internal logic to the external consequences of further system's outputs – for example, changes in economic practices "irritating" the legal system into producing new regulations and new laws "irritating" the economic system into altering its transactional behaviors – allows for a degree of coordination and mutual influence, where systems develop expectations about each other and react to changes in other systems.

5.2.2 The Cartesian Architecture Made Visible

Although a truly inspiring framework for understanding higher level governance cognitive systems as autopoietic, driven by internal logic and protocols, while indirectly sustaining the inherent *higher-order cognition* of such systems – by logically deriving that decision-making would emerge not from human level deliberation but from the system's own protocols, operational logic and architecture – as well as the affinities of its *structural coupling* framework with Active Inference plus how its functional definition of *system versus environment* tangentially resonates with the cognitive-light-cone and problem-space approach developed above³⁸⁸, Luhmann's framework nevertheless reinstates a profound dualism reminiscent of Descartes' separation between *res extensa* and *res cogitans*. That a system cannot *know* or directly interact with its environment in an unmediated way – i.e. that it internally constructs representations of the environment based on its internal protocols on models compatible with the logic it runs on – is naturally streamlined with the premises of both the Markov Blanket and Active Inference approach. However, where Descartes divided reality into material substance and thinking substance, Luhmann separates the world into *social systems* (operated autopoietically through communication) and *psychic systems* (human consciousness, which operates autopoietically through the medium of thought or consciousness itself), with humans awkwardly threading this divide as "structural couplings" between systems, instead of genuine participants, components or embodiments in either domain.

For Luhmann "*Humans cannot communicate; not even their brains can communicate; not even their conscious minds can communicate. Only communication can communicate*" (Luhmann, N., 1994a, p. 371), thus making humans parts of the environment of social systems (apart from the issues of coupling human embodiment with human thought, and human communication, which I will not get into as it falls outside the scope of this thesis, as well as any reference to conscious or the hard-problem³⁸⁹). Although humans are indispensable for communication to occur, for Luhmann, they are not part of the social system's autopoietic reproduction, given that psychic operations themselves are not considered elements within the social system. Consequently, the human as an embodied medium, or human intentions, motivations, subjective experiences, and psychological states while potentially triggering or affecting communication, are external to the direct, internal operations of social systems, which are governed by their own communicative codes and logics which remain impermeable.

The parallel to Cartesian dualism is striking. Like Descartes has separated *immaterial mind* from *material body*³⁹⁰, Luhmann created a chasm between human agents and higher

³⁸⁸ Chapter 2.

³⁸⁹ I would love to discuss the potential conscious state of higher-level cognitive systems, but this will have to be left for some other future endeavor. In total disclosure, I have no intuitions about this matter, nor have I pursued it yet.

³⁹⁰ As cited and explored at the first chapter

level governance assemblies, between *immaterial communication* and the *material humans* which communicate. His solution, "structural coupling", merely names the problem without resolving it, making it difficult to fully integrate the role of human beings as active constituents of the social systems they eventuate and enact. This thesis – which has previously, carefully and purposefully considered Cartesian dualism as an outcome to be avoided while selecting for its *reductionist substrate*³⁹¹ – does, due to my holistic and monistic commitment for an integrated understanding of reality, adopt an approach inspired by Luhmann's notion of autopoiesis, however it is greatly modified so as to accommodate the framework selected at the *Isms* chapter and the concepts showcased in chapter 2. From this monistic viewpoint, all phenomena, including the complexities of mental processes and the intricate dynamics of social structures, are ultimately rooted in matter-energy-force processes and the informational configurations they embody, be these departures from maximum-entropy on a probability space (Section 1.1), reductions in microstates compatible with higher-level assemblies, or reductions in generative models compatible with integrity (Section 2.1).

This Cartesian inheritance is not, however, a peculiarity of Luhmann's adaptation of autopoiesis. It is the same architectural inherence that Kelsen (1967) makes when separating norm from sociological fact. Luhmann's distinctive contribution is to articulate this architecture with biological insight and rigorous internal logic, making visible what other frameworks used to leave tacitly. The diagnosis therefore applies to the broader pattern, with Luhmann serving as its most analysable expression more than simply an *instantiation*.

5.2.3 The Empirical Costs

This dualist architecture generate analytical blind spots that handicap governance analysis at the level of explanation, prediction, and design, while artificially segregating Cognitive and Political science. Although nonexhaustively, three consequences are worth examining: the divorce of informational states across levels; the inability to account for assembly history; and the misanalysis of nested homeostasis.

Divorcing the informational states of lower-level sub-agents from higher-level systems results in the perceived separation between science – presumed to deal with discoverable truths about material reality – and the subjectivity of politics and governance, making as big a disservice to knowledge-seeking as separating mind and matter in neuroscience or psychology. Granting the mind's output autonomy from the neurons that compose it is the same as separating governance-systems from the characteristics and states of the humans and constituents that enact and co-compose them. Like Cartesian dualism plagued western thought, delaying whole fields of research³⁹², Luhmann then perpetuates a pervasive culturally shared intuition, in the form of a dualism which prevents/delays studying and explaining relationships between lower level human states and higher-level *Governance Cognitive Systems'* phenomena, such as: adherence or non-adherence to established constitutions, institutions and laws; the relationship between *cultural assembly indexes* and assembly dependencies on institutions or overall systems, and their adoption (i.e. which, and when, can certain *cultural-scaffolding* sustain given

³⁹¹ Chapter 1

³⁹² By locking *subjective experience* away from both methodology and empiricism, Descartes legacy rendered inaccessible the study of emotions, consciousness, pain... This separation of reason from the body's affective systems is the core of what António Dámasio deemed "*Descartes' Error*" – inspiring the title of this section – where he arguing that this legacy impoverished our understanding of the mind, while relegated non-human animals (and in this section, I argue, higher-level systems) to the status of mere automata.

governance models); thresholds in *cultural-assembly index* penetration (i.e. how many people share a given cultural-scaffolding) and higher-order maintenance or viability; higher level Governance Cognitive System's behaviour impact on lower level humans (a section on *disease* will be explored below); and many, many other questions relating to the human and cultural embodiment microstates' impact on Governance Systems and vice-versa³⁹³.

Furthermore, by externalising humans to the *environment* of the social system, Luhmann's framework doesn't accommodate the genesis of *Governance Cognitive Systems*, through the various first-principles operations described throughout this thesis (nor, curiously, does it account for the origins of communicative patterns themselves), turning higher level governance systems into impossible entities from an *assembly* perspective. While *structural coupling* – specially when viewed through the lens of Active Inference and lower level and higher level systems diminishing free-energy as per each-other's outcomes – is competent in describing a mechanism of coevolution and adaptation, it cannot account for the *assembly history* of the systems it sets to describe³⁹⁴, a negligence which sidelines the human bottleneck for its assembly, and thus the human embodiment and dependencies as part of the system itself (can chlorophyll be considered as part of the environment as opposed to the embodiment of a tree?).

The implications for understanding nested homeostasis and multi-scale competence in *Governance Cognitive Systems* are also impaired by externalising the human agent into the environment. When human agents dynamically pursue not only their own free-energy minimisation but also nested system-wide homeostasis, by seeking higher-level goals³⁹⁵ (instilled by cultural-scaffolding) in supra-individual functional tasks that promote higher-level integrity, are they simply “irritating” a social system? Underneath each *communicative* event there exists a myriad of nested and layered multi-agents, each processing information at their own level, each striving to maintain integrity and adapting to environmental influx. As the paper “Extraneous factors in judicial decisions” (Danziger et al., 2011) demonstrates, states of hunger or tiredness in a judge – manifestations of lower-level homeostatic drives down to the organelle – have deep correlations with which types of decisions is made, decisions which are functionally outcomes and parts of the higher-level legal system.

The same blind spots recur, with different vocabularies, across the broader landscape. Methodological individualism cannot explain how preferences are themselves constituted by the institutions whose effects they are meant to explain (Hodgson, 2007). Legal positivism cannot account for why officials accept the rule of recognition without circularity (Shapiro, 2011), because the question of acceptance is a question about embodied human states whose connection to the system the framework has axiomatically severed. Constitutional theory's debates over “*the constitution as written*” versus “*the constitution as practiced*” recapitulate at the doctrinal level the same dualism that Luhmann articulates at the systemic level, treating document and practice as separable and not as recursive phases of a single assembly process (Section 4.1.2). In each case the analytical move is identical, a level of the system is granted ontological autonomy by axiom, which then renders the dependencies that constitute that level's persistence invisible to the framework that posits it.

³⁹³ All these questions are those that the conditioning structure of Section 1.3.4 — where higher-level states *Y* reweight the lower-level distribution $P(X)$ into $P(X|Y)$ — were studied to address.

³⁹⁴ As per Section 1.2, the path-dependent accumulation of constraint configurations through assembly space, with the realised assembly graph heavily biased toward paths through robust intermediates by the path-reweighting selection mechanic of Section 1.3.4

³⁹⁵ A civil servant processing paperwork, a judge interpreting statutory language, or a citizen casting a vote — each performs operations whose full significance emerges only at the level of state-wide homeostasis and societal free-energy minimization.

5.2.4 The reckoning

Luhmann's main error then – by divorcing humans from the social systems – is failing to account that neither the hard protocols alone (which would be as useful as DNA without executioners) nor the dynamic drives alone (we have numerous examples – even today in functional failed states like Haiti – of individual human driven without hard protocols) can account for Governance Cognitive Systems. While both influence each-other enormously.

This diagnosis maps directly onto the architectural distinction developed in Section 2.6. Luhmann's framework, by treating social systems as constituted purely by communication while relegating humans to environmental status, theorises governance entirely through hard protocols while abstracting away the dynamic drives that any viable hard protocol architecture requires. The symmetric error, theorising governance entirely through individual rational choice while ignoring the systemic constraint architectures within which choices acquire meaning and within which preferences are themselves shaped, echoes through rational choice theory and methodological individualism. Both extremes fail because they treat as separable what is functionally integrated, i.e., a non-equilibrium system requires both the constraints that delimit its possibility space and the agents whose adaptive interpretations implement those constraints.

The cellular bioelectric network is not its voltage patterns alone, nor its individual cells alone. The planarian morphological assembly is not its bioelectric memory alone, nor its cellular dynamic drives alone (Section 2.2). By the same principle, no governance assembly is constituted by its constitutional code alone or by its citizens' actions alone. Frameworks that theorise either side independently produce, as a structural matter, accounts that cannot explain the persistence of the systems they purport to describe.

Despite the scientific revolutions that have since redefined cognition, embodiment, and systemic integration, many dominant frameworks within political science retain a tacit substrate selection that replicates Descartes' error³⁹⁶, presuming governance systems and political behaviour to emerge from abstracted ideals or institutional logics, divorced from the embodied, bio-cognitive, and material agents that enact them³⁹⁷. Here, I explicitly reject this dualist inheritance.

The framework developed across Parts I and II provide a monist account in which governance systems and the humans who enact them stand in constitutive relation, neither reducible to the other nor separable into autonomous domains. Constitutions are not blueprints external to populations any more than DNA is a blueprint external to cells (Section 4.1.2). Both encode constraint architectures whose viability is continuously produced and reproduced through the dynamic drives of the agents that interpret and enact them. Recognising this dissolves the scoping problem that has organised political thought, the question of where the system ends and the citizen begins, by reframing it as a problem of analytical level instead of ontological boundary. The cell is not external to the organism it constitutes, the citizen is not external to the governance system they constitute.

How cultural assembly indexes condition constitutional viability? How hard protocols and dynamic drives co-determine system persistence? How downward causation from

³⁹⁶ Nodding to António Damásio's title

³⁹⁷As detailed across Sections 5.2.1–5.2.3 of this chapter, particularly Section 5.2.2 "The Cartesian Architecture Made Visible," where I critically examine how this Cartesian residue subtly persists in Luhmann's theory of social systems — recasting mind/body dualism as system/agent dualism — and how this, when smuggled in as epistemological infrastructure, obstructs an integrative, monistic, and evolutionary understanding of Governance Cognitive Systems.

governance trajectories shapes the embodied substrate that those trajectories simultaneously depend upon? These are all lines of reasoning which open up and are worth pursuing.

Chapter 6 — Final Thoughts on Adding to the State of the Art

While chapter 5 diagnosed where I perceive the state of the art to fall short, this final chapter purpose is to make explicit what the framework developed in this thesis adds, that is, what becomes *thinkable*, *designable*, and *researchable* once governance is treated as a cognitive assembly continuous with the rest of nature. It is worth mentioning that this chapter remains descriptive and illustrative, providing non-exhaustive vectors for the ideas developed throughout the thesis.

6.1 What Becomes Visible

6.1.1 Governance morphology is governance action

It should first be recalled that the assembly en-route bottleneck towards Governance Systems and their *Constitutions* are individual humans with inherent preferred states (valence) who autonomously and dynamically pursue homeostatic goals. Through the dynamics illustrated in this thesis, individual human state-space is pruned yielding free-energy minimisation through lower individual and group entropy, leading to (via increasing cultural-scaffolding indexes) the co-creation, enactment and interpretation of higher level Governance Systems, which — when viable — eventuate stable post-singularity phases, with the joint distribution over agents concentrated on a strictly smaller alphabet than independence would allow, from which higher level goals and drives emerge.

The framework establishes that Governance Cognitive Systems do not act on their morphology as something separate from themselves, they act through morphological self-modification. Navigating own-code problem-spaces, so as to become sensitive to new variables and render new problem-spaces, isn't at all novel from a cognitive systems' perspective. Butterflies exemplify this phenomenon through their navigation of DNA expression problem-space during metamorphosis, reconfiguring their sensory apparatus to accommodate variables that are entirely absent during their caterpillar phase. During their larval stage, caterpillars operate within a constrained problem-space where sensory integration is primarily optimised for substrate-bound foraging³⁹⁸ and predator based triggered survival responses³⁹⁹ — utilising a sensory apparatus with limited spatial resolution and a sensory architecture that integrates a significantly more restricted set of environmental variables compared to the adult butterfly stage — which exhibits enhanced spatial navigation

³⁹⁸ Caterpillars (this study focused on silkworms) have an ultra-specific, narrow and individually based chemical "authentication system" to determine if it should feed on or not on a given plant. Using a mandatory "triple lock" recognition system requiring simultaneous detection of three specific mulberry compounds: chlorogenic acid (CGA), isoquercetin (ISQ), and β -sitosterol. "If these three compounds are detected simultaneously, the larva makes a test bite, which leads to the second step" (Endo et al., 2024)

³⁹⁹ "In this study, we demonstrate that the hairs of *Orgyia leucostigma* (Lymantriidae) caterpillars not only function as physical barriers that deter predators but also act to provide sensory inputs triggering behavior that reduces predation risk. In particular, the way in which caterpillars respond when their hairs are touched is predator specific. Mechanical parameters of the interaction determining the response were identified correlatively and confirmed manipulatively. Caterpillars predominantly dropped in response to high hair bending velocities and predominantly walked away in response to low hair bending velocities. These stimulus-specific responses appear to be adaptive as they led to increased survival. Our results demonstrate a functional link between morphology and behavior." (Castellanos et al., 2011)

systems⁴⁰⁰ and three-dimensional flight-oriented sensory integration⁴⁰¹. The difference and novelty is that – while the caterpillar's genome already contains the complete cascade of triggers and solutions whereby butterflies' sensory systems eventuate – Governance Systems encode processes to generate, integrate and codify whole new code solutions.

Whereas a butterfly's metamorphosis runs a developmental script already encoded in its genome, Governance Systems encode processes to generate whole new code, yielding new senses, new organs, new variables they become capable of measuring and acting upon. For example, the German (Grundgesetz - GG) and Portuguese (Constituição da República Portuguesa - CRP) *Constitutions*⁴⁰² – mostly through Articles 76-78 GG and 167-169 plus 134-136 CRP – explicitly encode code generative pathways as meta-instructions for creating new "*Hard Protocols*". An example of the German Governance Cognitive System navigating its own morphogenetic code as a problem-space so as to generate an organ sensitive to new variables, can be illustrated with the 1974 Law on the Establishment of a Federal Environment Agency creating the Umweltbundesamt (UBA)⁴⁰³. This allows the system to become sensitive to a whole new class of quantifiable variables, opening the door to — for example — the ambient concentration of Nitrogen Dioxide (NO_2) as variable. These new variables open up emission zones problem-spaces and “ecosystem health” goals (amongst others) previously unattainable. A similar expansion in Portugal's *Cognitive Light Cone* occurred when, through Law No. 18/2003, a Competition Authority (Autoridade da Concorrência, AdC)⁴⁰⁴ was created, to open up new variables on a system once focused mainly on less granular macroeconomic variables like GDP, so as to enable anti-market-distortion goals, becoming sensitised to variables such as market share or the Herfindahl-Hirschman Index. The ability of both systems to measure and act upon these variables expanded their *Cognitive Light Cone* and revealed new problem-spaces to further navigate. In both instances, the Governance Cognitive System deliberately approached its own morphology as a navigable problem-space – rewriting or extending its own code by running generative code protocols – to build new senses, thereby defining new actionable spaces to interact with. A change in interest rate from 2% to 2.5% is a change in physiology, the introduction of a central bank is a change in morphology, both being governance actions. Only through the framework developed in this thesis renders them legible as continuous along a single dimension of cognitive self-modification.

6.1.2 A non-ideological taxonomy of Governance Systems

This morphology-as-action recognition immediately yields a new taxonomy for governance systems beyond traditional qualitative ideological labels. Instead of classifying systems as democratic versus authoritarian, capitalist versus socialist, or liberal versus illiberal, the framework permits classification by their morphology, the breadth and depth of

⁴⁰⁰ “the butterflies maintained robust goal directions relative to a virtual sun. By resetting their goal directions, we found neurons whose spatial tuning was tightly linked to the goal directions. Importantly, their tuning was unaffected when the butterflies changed their heading after compass perturbations, showing that these neurons specifically encode the goal direction” (Beetz et al., 2023)

⁴⁰¹ “Directional visual information are sent to the central complex, a brain structure involved in sensory-motor integration and locomotor control in the context of processed internal and external stimuli, and the likely site of integration of both the circadian clock and skylight information, important components of navigational control” (Couto et al., 2020).

⁴⁰² As defined at the subsection Section 4.1 (Stable recipes for iteration)

⁴⁰³ As explained in the following website: 50 years German Environment Agency: A retrospective. Retrieved August 11, 2025, from <https://stories.umweltbundesamt.de/en/50-years-german-environment-agency>

⁴⁰⁴ The translated legal document can be consulted here: <https://www.wipo.int/wipolex/en/legislation/details/7897>

their *Cognitive Light Cone*, and the coordinates opened up in the problem-spaces they can inhibit. Two systems sharing an *ideological* label may differ vastly in the variables they render, the organs through which they sense and act, the assembly index of their constitutional code, and the structure of their joint distribution over component states (Section 1.1.3). On the other hand, systems labelled antithetically may share more cognitive architecture than either tradition acknowledges. The taxonomic question shifts from "what does this system stand for?" to "what can this system see, model, and act upon?".

6.1.3 The dependency on lower-level dynamic drives

Regardless of acting in lower-level embodiment problem-spaces, the goals motivating Governance Cognitive Systems' operations within them are still higher-level ones. For example, when the CRP (Article 65^o) establishes the right to housing, its goal is not to satisfy and directly address a specific individual's housing conditions, but rather of having a housed population and available to the population the mechanisms whereby the state enables supports and enacts a drive towards satisfying the metric "housed population". Having a high housing index drive is as remote to an individual human as having a high GDP per capita, two variables that although tapping directly into human drives, are abstracted from individual instances and reflect higher order societal goals, irreducible to a single individual. Regardless of being able to directly approach its morphology and code to ensure organs, offices and state functions⁴⁰⁵ to address housing, the system's goals are upheld and determined as higher-level abstractions derived from homelessness data, housing market indicators and such abstractions, however derived from individual lower level drives. Despite not navigating towards lower-level agent human needs directly (rather towards higher-level goals abstracted from these), Governance Cognitive Systems find in their lower-level constituent human makeup – as part of their embodiment – a problem-space to navigate towards their higher-level goals.

As a result, whole classes of increasingly higher-level goals are generated, very often highly abstracted from the initial lower level agentic needs. However – as the assembly index increases – there is nothing stopping higher-order entities from tapping into lower level physiological or metabolic goals directly, so as to satisfy them from a system's level route strategy, previously unavailable to the lower level agency. Picture whatever science fiction scenario in which future humans come to inhabit extreme water scarce environments, and choose to edit their genetic makeup so as to create organs and metabolic solutions that perfectly recycle water so as to make urination and sweating unnecessary. This case of system level architecture directly solving at lower level, however to solve a higher-level integrity drive, exemplifies the kind of everyday operation that takes place at Governance Cognitive System's problem-spaces.

As the system's embodiment is both the code and the agents (plus the whole of the mediums in which the system runs), and given that the code's manifestation oscillates in "genotype vs phenotype" like dynamics, the Dynamic Drives and cultural-scaffolding of lower agentic level mechanics interpret and co-create the Hard Procol code differently and in a context dependent manner, illustrating the same underlying dynamics of multi-agentic systems relying on bureaucracy for integration and viability, as illustrated in Section 2.5. Many of the higher level actions of Governance Cognitive Systems lean directly into the multiscale competence of its lower level constituents, acting in human behaviour (motivators

⁴⁰⁵ Functional roles abstracted from individuals, as showcased above at the Governance Cognitive Systems in Assembly Index

and deterrants) problem-spaces so as to influence the decision-space of said agents, constrict their state-space to coordinates compatible with the higher level desired states.

6.1.4 Policy as morphological active inference

Once the three preceding points are taken together, governance navigation can be conceptualised. A Governance Cognitive System's location in problem-space is defined by the joint configuration of valences across all variables it is sensitive to. Its sensitivity is determined by its morphology — the organs, registries, and instruments it has assembled. The system "moves" through problem-space by changing its own morphology and the speed at which it can move is set by the speed at which its *Constitution* can be iterated. As the assembly index of the *Constitution* increases, the dimensionality of the problem-space the system can render expands accordingly, with each new variable integrated, it opens new coordinates that were previously beyond invisible, that is, nonexistent from the system's perspective.

This yields a dual-optimisation structure native to all viable cognitive systems and now made explicit at governance scale. The system simultaneously pursues, first, the search for preferable coordinates within the problem-space its current morphology renders; second, the meta-search for a more complete problem-space, achieved by integrating new variables and the organs that render them. The first is policy while the second is constitutional iteration. Conventional political analysis treats these as separate domains, ordinary law-making versus constitutional change, while this framework reveals them as two timescales of the same operation of morphological active inference, performed on a problem-space whose dimensionality is itself a function of morphology⁴⁰⁶.

Local optima that look successful within the system's current variable set may be artifacts of poor sensing or under-modelling. A policy that yields high GDP and low inflation but low Human Development Index will not yield error signals from the system's primary economic sensors, yet may sit on what is in fact a depressed coordinate once HDI is added to the model. The drive to reconcile the generative model with the ontological world — to minimise the model's overall uncertainty, not merely the distance to its preferred coordinates — is therefore not a design preference but a structural requirement of any non-equilibrium cognitive system that wishes to remain viable. Viable Constitutions encounter selective pressure⁴⁰⁷ to engage in actions whose primary value is epistemic, with census, registries, audits, statistical agencies, scientific institutions...

Beyond errors of sensing, even a faithfully modelled landscape can confront the system with the structurally harder problem of trajectories toward higher-valence coordinates that are reachable only by passing through a stretch of the problem-space where every intermediate reading degrades. What might be termed a *valley of sacrifice* is the topological feature in which the local gradient and the global gradient point in opposite directions, in that each immediate step registers a worsening across the variables the system is most acutely sensitised to, even though the destination on the far side dominates the origin on those same variables. The system cannot reach the higher coordinate by monotonically improving each intermediate reading, and must instead act against the proximate signal in service of a

⁴⁰⁶ The system navigating toward configurations at which its constraint family undergoes a non-analyticity (a phase transition in the sense of Section 1.3.2) or its compression-relevance Lagrangian undergoes a bifurcation (Tishby, Pereira, & Bialek, 1999), thereby crossing a singularity-node (Section 1.3.3) into a new post-transition phase whose macrostate alphabet was inaccessible before.

⁴⁰⁷ in the precise statistical sense of the path-reweighting mechanic of Section 1.3.4, where the realised distribution over governance assemblies departs from the prior distribution in proportion to the persistence and propagation of each path's terminal configuration

distal one its current sensors cannot yet register as gain. Portugal's 2011–2014 trajectory under the Troika adjustment programme (Dias et al., 2014) illustrates this dynamic at governance scale. Fiscal consolidation, labour market reform, and structural adjustment registered as deterioration across nearly every variable the Portuguese governance system rendered with high sensitivity, from rising unemployment to real wages falling, GDP contracting, public service capacity diminished, social discontent surged... Yet the trajectory was a (contested but identifiable) traversal toward coordinates — restored sovereign borrowing capacity, recapitalised banking sector, renewed external credibility — unreachable by any path along which each intermediate reading improved.

Valley traversal exposes a structural alignment problem between the temporal scope of a system's generative model and the temporal scope of its feedback architecture. Electoral cycles, news cycles, and the attentional bandwidth of constituent populations are most responsive at the timescales over which the valley appears as deterioration, generating dynamic-drive pressure (Section 2.6) for course reversal at every step of the descent. A system whose hard protocols cannot hold against this pressure, whose long-horizon commitments remain revisable on short-horizon attentional swings, will systematically prefer accessible coordinates over distal ones even when the latter are correctly modelled and known. Conversely, systems that successfully traverse *valleys* deploy what is, structurally, assembly memory at the constitutional scale, with prior traversals crystallised as institutional precedent and narrative scaffolding, hard-protocol provisions that protect long-horizon commitments from short-horizon revision, and metacognitive layers (Section 6.2.3) capable of overriding the immediate gradient in service of a generative model whose horizon includes the far rim. Each successful traversal becomes a reusable assembly bottleneck assembly bottleneck (Section 1.2.1), narrowing the search space for subsequent traversals by encoding "this kind of valley can be crossed, and crossed thus" into the system's cumulative memory. The capacity to cross is therefore not a property of any single policy decision, but of whether the system's cognitive light cone is wide enough to encompass the far rim as a coordinate in the same model that registers the trough.

The system's trajectory is thus not a search across a static landscape, but the very process of *drawing* that landscape, thus a recursive cycle of confronting ontological surprise, integrating error through morphological self-modification and broadening sensitivity through variable integration. This is the core mechanism by which a Governance Cognitive System world-model can be sustained. A system performing this recursive cycle needs structures that perform feedback integration, memory consolidation, attention and metacognition, and cross-level model coordination, as the operations any cognitive system performing the dual optimisation above must perform. Their absence or non explicit design so as to facilitate it, in today's governance architectures is — from this thesis vantage point — a cognitive deficit which out to shift from "by default" to "by design" in its optimisation and intentionality.

6.2 Hard-Protocol Design Implications

6.2.1 Feedback integration

One way in which the architecture of Hard-Protocols could be leveraged functionally so as to enhance the Governance System is by designing explicit feedback integration mechanisms whereby the world model and corresponding actions are calibrated. What is at stake here can be sharpened against a default that Cyert and March (1963) documented at the firm scale, and that, left unchecked, will reassert itself at the governance scale, where

firms *avoid* uncertainty instead of building predictive apparatus to navigate it. They prefer, in the authors' words, "*avoiding planning where plans depend on predictions of uncertain future events*" and "*emphasizing planning where the plans can be made self-confirming through some control device*"⁴⁰⁸, i.e., negotiated contracts, buffer inventories, financial reserves... The result is short-run reactivity in place of long-run anticipation. Governance Cognitive Systems are not exempt from this default. Without explicit architecture pulling against it, hard protocols will tend to ossify around self-confirming control devices (negotiated revenues, fiscal reserves, electoral horizons that match attention cycles) without developing the predictive apparatus that long-horizon problem spaces require. Making explicit feedback integration architectural is therefore not optional refinement but a structural countermeasure against an attractor the framework itself predicts. Either by incorporating into individual laws, bills and regulations themselves mechanisms whereby its application, execution and goal distance amongst (other potential feedback variables) is checked; by creating a feedback module responsible for monitoring and calculating the difference between desired outcomes and actual consequences per domain or goal; or by any other architectural solution that makes it so that each Governance action has clear explicit goal parameters and success metrics to be matched against feedback, and a feedback protocol to control and calibrate it⁴⁰⁹.

6.2.2 Memory consolidation

Memory consolidation — and the progressive demarcation between working, short and long term memory — could also be leveraged so as to have more adaptive, resilient and context dependent behaviours at the Governance level. Legal systems' tendency to become bloated and overly complex have various culprits, but a sure one is the absence of expiry or elapse date or mechanism in the great majority of its actions⁴¹⁰. An architectural solution whereby actions, their feedback and success metrics are stored in relation to specific goals and scenarios, and stored in relation to its recall and recall success, so as to mimic the mechanisms underlying memory, would make it so that hyper-specific actions could be taken in a nimble manner which would make them more niche and deployable, acquiring a more heightened calibration or weighting the more it was deployed as per its degree of success. Thus a landscape of concurrent solutions competing for circumstantial fitness could coexist in a memory of more readily to less readily available actions to take, as per the concrete demands of the situation and the hierarchy of values and goals in place. Local fire solutions in a geographic location comprehended within the same Governance border could be non-binding for other geographic states until its usage and success matrix under certain variables passed a threshold that would make it obligatory for deployment across the territory under the same variable landscape.

This would allow short term localization of solutions and experimentation, validating and testing hypotheses hyper-specific to the relevant metrics, that could be recalled and deployed at other locations as its memory gets *deeper* and with higher weightings. Iteration on this governance functional module could lead to flexible and context dependent policy

⁴⁰⁸"[Firms] avoid the requirement that they correctly anticipate events in the distant future by using decision rules emphasizing short-run reaction to short-run feedback rather than anticipation of long-run uncertain events. They solve pressing problems rather than develop long-run strategies" (Cyert & March, 1963, Chapter 6).

⁴⁰⁹ Operationalising, at the governance scale, the prediction-error minimisation loop of active inference (Section 2.1) and provide the feedback channels through which the path-reweighting selection mechanic of Section 1.3.4 can act on policy paths in proportion to the persistence of their outcomes

⁴¹⁰ This is the failure to distinguish, within governance assembly memory (Section 1.2), between configurations that have proven viable across multiple deployments (Section 1.2.1) and configurations whose terminal effects have decayed or whose problem-space coordinates have shifted out of range

scripts, deployable across dynamic situations, as opposed to a fixed if-then legislation. Governance action is famed for being reactive (as opposed to pro-active), given current governance structures tend to resist change until collapse⁴¹¹. Having lower level governance markov blankets experimenting with legislation for specific parameters, and having it weighted against other such similar solutions for close parameter coordinates, allows for the organic development of local scripts to be implemented as needed.

6.2.3 Attention and metacognition

A further way in which *Hard-Protocol* architecture can be explicitly designed so as to eventuate classic cognition modalities, is towards achieving or perfecting system level attention mechanisms. Attention in today's governance cognitive systems is captured by default (as opposed to by virtue) through the news cycles⁴¹², "*Policy entrepreneurs*" engaged in narrative shaping to promote a frames, perspectives or stories to convert events into attention, to drive public opinion and state action momentum⁴¹³ and "*public mood, partisan composition (...), and the organization of interests around issues*" (Johnson, 2006). Having a structure akin in function to how Karl Friston (Sandved-Smith et al., 2021)⁴¹⁴ envisions attention and meta-cognition — by layering "checkers" controlling what are the processes the previous layer is attending to — built into the Governance architecture, would enable the System to have an own level mechanism whereby to understand where attention is, where it ought to be in relation to goals, priorities and subgoals, and understanding the opportunity cost of paying attention to one thing as opposed to another.

This can then be further leveraged so as to introduce cognitive phenomena that require meta-cognition, such as — for example — delayed gratification, by allowing the system to divorce or recalibrate from the immediate and volatile lower level attentional driving forces. Creating a control layer nested under a "sustainability" or any other long-term goal, and having it check if the current information processing status is compatible with it while redirecting or adapting the current status to something that is, is a higher level concept that is worth exploring and trying out in a Governance lab. This would not only aid in agenda setting and issue prioritisation, but could harmonise execution, methodology and goal pursuit to align and keep being redirected to higher weighted goals.

6.2.4 Cross-Markov-blanket generative model coordination

Section 6.1.4 framed governance navigation as morphological active inference performed on a problem-space whose dimensionality grows with assembly index. At any given moment, however, governance is not a single inference engine but a nested

⁴¹¹ First a disaster happens, like the 2008 market collapse, and only then legislation is passed informed on the disaster to mitigate the changes of future recurrence.

⁴¹² "Another consequence for governance processes has to do with how public actors react and are organized. Because public actors are encountering more media attention and this media attention is important for the development and result of governance processes, as several authors argue, more and more communication and media professionals are to be found within governmental institutions. This has gone hand-in-hand with the use of marketing tools and brands in political and public communication and in the presentation of policy plans. The communication professionals must cope with media logic in their pursuit of as much positive publicity as possible, as well as in their attempt to protect their organizations against negative or undesirable publicity" (Klijn & Korthagen, 2018)

⁴¹³ "political entrepreneurs (...) engage in an effort to reframe the issue. This activity led to changes in media frames because reporters pay close attention to such political entrepreneurs. Drawing on a content analysis of the ABC and CBS evening news, Callaghan and Schnell show that shifts in media frames of gun control closely track with the activities of policy entrepreneurs." (Johnson, 2006)

⁴¹⁴ Besides what is already stated at Section 2.5

architecture of Markov blankets — individual, familiar, municipal, regional, national, supranational — each operating its own generative model.

There could be implemented a modality across these governance Markov blankets that operates as generative-model builders, determining the variables, metrics, collecting feedback, sampling the constituents' hierarchy of values, and producing short, mid and long term forecasts of the model's coordinates. By providing information and receiving information across levels, it would promote surprise pruning and challenge poor signaling. The better or worse performance of lower level governance Markov blankets could be informative in calibrating which signals are sent to the higher levels with more or less weighting, or inform auditors on what and how parameters are better tested, integrated or deployed.

6.2.5 A broader claim

In sum, I would just like to stress through these surface level — to be developed and non-exhaustive — examples above, that understanding cognition through a functionalist lens, allows us to approach Governance Cognitive System's Hard Protocol design with cognitive characteristics in mind that — having long been understood as properties in other cognitive systems — can be deliberately inscribed into the informational processing protocols of higher level cognitive systems. Understanding the system's state-spaces compatible with both integrity complemented with further cognitive characteristics, it becomes possible to embed and experiment within governance different protocols to promote higher-level prediction, adaptation, goal alignment and so on. Again, this thesis is not normative nor prescriptive, thus the examples above are mere illustrations of my broader and most critical point, that cognition is instantiated in maintaining lower entropy and integrity in non-equilibria systems⁴¹⁵, that strategies to achieve these macro goal have dependencies on nested subgoals which are operationalised via a myriad of different solutions, these can vary from less to more effective, optimise towards vastly different valence landscapes and hierarchies or preferences, yet their instantiation runs on an architecture of information processing that can be a matter of study across scales. The modalities of uncertainty reduction and the operations of constraint imposition should be studied as such across substrates and levels. The point is far from being that Governance Cognitive Systems should think like brains, but that the principles underlying information processing across systems can be understood, inspired and derived so as to inform higher-level societal system design. By recognising this, academia and social representations of governance can evolve so as to render Governance System's cognitive morphology explicit, opening the door to deploying methodologies, vocabularies and paradigms which are critically missed in policy-making, because we haven't taken a monist approach to cognition yet.

6.3 Dynamic-Drive Design Implications

In the same spirit of the exercise above — as multiple Hard Protocol architectures lead to wildly different Governance System's cognition types and modalities, which can be leveraged by leaning into, studying and understanding principles across cognitive disciplines — so too does understanding and leaning into cultural-scaffolding and human Dynamic

⁴¹⁵ Characterised in any of the three complementary registers developed in Section 1.1.2 (negentropy of single distributions, mutual information across coupled variables, total correlation among the components of a system at one level)

Drives, as understood through the multiscale competency paradigm, open the door to new understandings and novel public policy designs.

If Section 6.2 addressed the system's "genotype" — the codified protocols — this section addresses its "phenotype", that is, the dynamic, situated, motivated interpretation through which any Hard Protocol becomes embodied governance action.

6.3.1 Policy that leverages multiscale competence

As previously detailed, any given Hard Protocol is eventuated and materialised through the dynamic and context dependent interpretation of its agents, in line with the dynamics highlighted at Section 2.5, in which the same code can instantiate markedly different system-level state-spaces depending on the contextual and motivational milieu of its executors. Hard Protocols aren't fixed operators in abstract, but a trajectory from specification to local interpretation, embodied enactment and finally feedback, with the gap between specification and enactment being in itself a nested competent cognitive problem-space.

Understanding this means realising that public policy should be treated as a co-designed interface between higher-level intention and interpretive agent drive, in the sense that clauses should carry minimal executable hard-lines anchored in contextual information regarding the intentions and desired outcomes at the higher level. These should be designed with likely frontline interpretations in mind, in line with institutional memory of previous outcomes. On the long run, situational interpretations are bound to become context-indexed, and legislating can go beyond frontline rule fixing to become feedback curating and pruning of interpretative behaviour alignment with protocol priors and intentions, in a context dependent manner. Taking advantage of the multiscale competence of human agents, policy can therefore be deliberately ample so as to have room to lean into situational interpretation — aligning individual dynamic drives with system level desired results, while relying on situational feedback and auditing — allowing a dynamic, plastic and context dependent deployment of competent agents aligning dynamic drives with higher level interests, instead of a set of imposed fixed hard protocols which sacrifice nuance, context and fitness for standardisation and a notion of justice that treats what is different the same.⁴¹⁶

A contemporary piece of legislation, which makes a stellar example illustrating the tip of the iceberg in the potential for this type of policy design, is *The Money Laundering, Terrorist Financing and Transfer of Funds Regulations 2017 (UK)*. Given that fraud prevention and anti-money laundering is a cat-and-mouse game (between the legislator, financial institutions and fraudsters/criminals), any fixed hard-coded rules stipulating specific controls, prohibitions or alert types are bound to become obsolete as the bad agents find novel strategies and patterns to structure assets, defraud individuals, place and integrate funds. Aware of the futility of navigating the whole legislative process incrementally to add new and never-ending red-tape on the back of novel money laundering and fraud types and patterns, the legislator has decided to lean into individual human incentives and competence. This is achieved by creating a new Hard Protocol functional and abstracted role which nevertheless is designed so as to nest within it the multiscale competence of the individual human-level agent embodying the role, by creating the Money Laundering

⁴¹⁶ Alining close to the Aristotelian definition of justice, of treating equals equally and unequals unequally in proportion to their relevant inequalities, but applied at the system's level and beyond the scope of individual rights and punishments.

Reporting Officer (MLRO) role. Financial Institutions (FI) are required to hire and abide by the MLROs' compliance stipulations, while these (MLROs) are personally and criminally liable for infringements committed by the overall FI. This is particularly useful given that what constitutes an infringement is a situated and context dependent negotiation with the regulator through its anti-financial crime bodies and auditors.

As the legislation follows a Risk-Based Approach — as opposed to hard rules fixing methodology — it is up to the MLRO to understand and proactively mitigate the FI's specific risks, as per its demographics, product line, history and overall characteristics profile. Ongoing audits and conversations with the regulator — as well as a comprehensive Suspicious Activity Report / Disclosure methodology in which the FI is liable for all the suspicious activity it doesn't disclose — make it so that risks can be discussed and aligned, presented and warned, in a manner in which the personal liability of the MLRO is tempered with its proactiveness, effectiveness and alignment with the higher-level systemic intentions of fighting financial crime. By leveraging individual responsibility and interests, with feedback and situational risk assessment, the higher-level Governance system adds dynamically to its toolkit a myriad of context dependent solutions that would never be reached in a fixed-rules hard red-lines scenario, by acknowledging the overall multiscale competence of its architecture.

6.3.2 Surfacing latent drives

Understanding multiscale competency, and the phenotype-like *interpretation to realisation* gap in implementing Hard Protocols, is crucial not only to design specific policies such as the one illustrated above, but more fundamentally to be incorporated into the core of the hard-protocols themselves. This is crucial since tensions arise — both at the decompression of bowtie bottlenecks and at the more general application of Hard Protocols — due to the ever present need for interpretation which is always situated within motivational parameters and nested within multiplicities or values hierarchies. While it is acknowledged the range in perspectives on the role individuals should/can have on the interpretation of the Constitution itself, much less often are discussed design principles that could be established so as to take into consideration the wide range of dynamic-drives motivating hard-protocol alignment and interpretation. To illustrate the traditional manner in which *interpretation* on a narrow sense is discussed in Political science, allow me to exemplify with the tension between, on one side, *popular constitutionalists* — as defined in the book “The People Themselves” by Larry Kramer (2004)⁴¹⁷ — and positivists. Popular constitutionalists defend “*judicial review without judicial supremacy*”, “*taking the Constitution away from the courts*”, thereby making an explicit “*distinction between the ideas of judicial review and judicial supremacy*”, thus opposing that “*judges have the last word when it comes to judges have the last word when it comes to constitutional interpretation and that their decisions determine the meaning of the Constitution for everyone*”, that is, that “*courts should not have “normative priority in the conversation” about the meaning of the US Constitution*”, and given that branches of Government function in a dynamic of checks and balances “*none of the branches of government could claim superiority over the others, in interpreting the Constitution*”⁴¹⁸.

Larry Kramer thus “*considers that courts— and particularly the Supreme Court— should start recognizing the people as the higher (interpretative) authority*”. On the other side

⁴¹⁷ As cited by (Sellers & Kirste, 2023)

⁴¹⁸ Quotes in this paragraph from (Sellers & Kirste, 2023) citing (Kramer 2004a, 125)

of this tension representing a positivist view on legal interpretation, “*Alexander and Schauer (1997) (...) argued that judicial supremacy in constitutional interpretation is necessary to avoid interpretive anarchy*” (Sellers & Kirste, 2023). However, even if a compromise was reached dictating a referendum on the interpretation of the Constitution under certain circumstances, the manner in which this rule could be encoded would still be subjected to interpretative forces that would influence when and how such a referendum would be held regardless of how exact the letter of the law. Not only do terms evolve, but the relationship of the culture, niches, and people evolve with the language. What truly matters is the type of relationships underpinning how the letter of the law is expressed. Underneath there is a meta-law of what are we agreeing language to mean (as seen recently and with sensationalism on how whole legal problematisation around the meaning of the word “woman”, which has been shifting in the collective understanding of what it ought to represent).

A more useful way of uncovering the dynamic drives underpinning motivated interpretations across Governance Cognitive Systems' processes and domains, is by making explicit the range of factors which contribute to having a set of interpretative tendencies over others⁴¹⁹. Having constituents vote on the hierarchy of values for the state to pursue, attributing representation to proportions of personality types, ages, priority typologies and time-span concerns would enrich representative democracies with explicit sets of preferences with more granularity and predictability than mere party filiation, while freeing the Governance Cognitive System to pursue novel solutions aligned with the subliminal core motivators of interest, without having to address constituent concerns by proxy of the surface-level problem being addressed.

Uncovering the dynamic drives underpinning motivated interpretations across Governance Cognitive Systems' processes and domains is to surface, measure, and bind the latent variables that actually steer interpretation. Assuming that nested within the dynamic drives are relatively stable (idiosyncratic and cultural), but context-sensitive cognitive–affective propensities that bias attention, valence and cognitive style (e.g., threat/purity sensitivity, novelty-seeking vs. risk aversion, fairness/care weighting, temporal horizon of concern), thus informing preferences and so interpretation. These drives constitute the priors that drive the phenotype-like gap between specification, local interpretation, embodied enactment and system feedback at scale within the Governance System. Making these explicit — as opposed to letting them implicitly govern how Hard Protocols are deployed and perceived — allows the system to co-design policy as an interface between higher-level intention and frontline cognition without having them be a one-way imposition of rules.

As stated above, having constituents vote on the hierarchy of values should be understood as meta-policy to informed policy design. A periodic, domain-specific statement of goal priorities (e.g. safety, liberty, equality of opportunity, innovation capacity, ecological integrity), expressed as a rank or weight vector. Likewise, time-indexing (near-term, mid-term, long-term) is fundamental (especially as the developed world becomes a gerontocracy) because the citizens endorse different trade-offs across different time horizons and have different values and attitudes as per their life stage, and Governance Cognitive Systems must arbitrate explicitly temporal conflicts with drives privileging some life stages over others. This means moving beyond purely territorial or partisan representation to

⁴¹⁹ This is to make legible the conditioning structure that reweights the realised distribution of interpretations away from the prior distribution over possible interpretations, in proportion to the persistence and propagation of each interpretive trajectory

statistical representation of cognitive heterogeneity, to ensure that the distribution of drives (e.g., high-vs-low risk tolerance; near-vs-far time orientation; body-salient vs. mind-salient framing) is actually present in deliberative venues and explicitly encoded in the objective functions used by agencies.

These become (i) constitutional intention priors for agenda-setting and auditing; (ii) northern stars from which to optimise interpretation towards; and (iii) sampling frames for deliberative panels, potential sandboxes, and explicit root-cause discussion. For example, consider immigration adjudication, where a ballot is run with three time horizons from which to assign weighting (0–2y, 2–10y, 10–30y), then a section where constituents rank values such as (D) diversity, (S) public safety, (L) rule-of-law integrity, (E) economic dynamism, (H) humanitarian protection, (C) cultural continuity and (I) innovation capacity, and finally a questionnaire is developed to understand potential cognitive style (change averse vs novelty seeking, empathetic vs individualist) corroborates. With such explicit information, the Governance Cognitive System knows within which parameters to act so as to address not only proxy stereotypical positions on beyond core concern issues, but the explicit content of constituents' concerns, the structural profile of the drives and trade-offs that those concerns emerge from.

This transforms governance from a reactive stance, where institutions respond to surface-level issue framings, to a generative stance, in which the system can design interventions, sequencing, and framing strategies that pre-align with the motivational topography of the polity. By operationalising constituent drives and priorities in this way, the Governance Cognitive System gains a transparent, auditable bridge between higher-level constitutional goals and the situated, embodied cognition of frontline agents. It also creates a formal mechanism for integrating cognitive style (such as highly future-oriented innovation advocates in an otherwise present-focused, benefit-seeking electorate) into system-level decision-making, without over-representing them. Over time, this design fosters a meta-awareness within the polity of its own cognitive–motivational architecture, allowing democratic deliberation to occur not just over outcomes, but over the modes of cognition that produce them.

Furthermore, as AI will inevitably play a more and more integrated role in policy making and governance, staying closely anchored to human-level preferences and concerns at an explicit and granular level will be critical.

6.4 Future Research

This thesis opens lines of investigation that were structurally unaskable within either Cognitive or Political Science as currently positioned. The following sections are a non-exhaustive draft of questions I find most pressing and interesting, which ought to be read as one continuous research programme broken into four overlapping vantage points⁴²⁰, which become pursuable once Governance Cognitive Systems become conceptualised.

6.4.1 For Cognitive Science

The criteria for systemhood, level-consolidation, and selection⁴²¹ are checkable in principle against the structure of the relevant joint distributions, plus the apparatus exists in statistics and information theory to check them in practice. The question of when a collection of components becomes a system can be operationalised in any context where the joint

⁴²⁰ A reader by now ought to be familiar with my aversion to siloed disciplines

⁴²¹ Developed in Chapter 1

distribution of the components admits *estimation*. Neural ensembles, cellular collectives, multi-agent human teams, and institutional architectures are all candidates. I would like to see it deployed even if alongside the variance-based and correlation-based measures that have historically dominated, particularly in contexts where the dependence among components is non-linear or where the sign of pairwise covariance varies across pairs.

A related opportunity concerns the identification of *singularity-nodes*. The information-bottleneck Lagrangian, in particular, provides a parameter-free way to locate the configurations at which a joint distribution undergoes a *qualitative reorganisation*. Applied to developmental trajectories, learning curves, or cultural-historical transitions, this offers a principled alternative to the threshold-based criteria. Where does a child's joint distribution over linguistic states bifurcate into the macrostate *speaker of language L*? At what point in cultural assembly does the joint distribution over communicative practices undergo the bifurcation that produces *literate population*? These are questions the framework lets us pose.

An equally important shift this framework enables is tapping into bowtie compression, adaptive interpretation, nested Markov blankets, and active inference as recurrences of the same informational architecture across substrates. A planarian fragment regenerating from compressed bioelectric pattern memory and a regulator enforcing statutory provisions from compressed documentary memory are, in this view, two cases of the same problem, and their differences are informative precisely because the underlying mechanic is shared. What follows is a methodological implication given that the framework dissolves the individual/distributed dichotomy by treating both as coarse-grainings of the same probability space. I think much of the heat in extended cognition debates over the last twenty-five years comes from treating the question of "*where cognition really lives*" when it should be treated as a question about the appropriate level of analysis for a given inquiry, given the cognitive light cone of the system at hand. Once one accepts that the higher-level probability content follows by pushforward from the level below, the question stops being whether cognition extends and starts being which coarse-graining the inquiry's question requires.

The framework also predicts that an underlying, assembly *selection mechanic* of path-reweighting⁴²² operates at every scale where viable configurations persist and propagate. If this is right, then evolutionary biology, cultural evolution, technological innovation and institutional evolution should all admit description in a common assembly-graph vocabulary, with their respective dynamics differentiated by the substrate-specific properties of their bottlenecks rather than by separate first principles.

The final research line for Cognitive Science is the one I think most needs to happen. The three frameworks this thesis has leaned on most heavily (Basal Cognition, Assembly Theory, and Active Inference) have each on their own accumulated enough empirical and formal weight to be taken as constitutive contributions to the field. Cognitive Science has not yet integrated them fully and in an articulated manner, and I believe it should. Read together, the three converge on a definition the discipline should adopt, that any assembly actively maintaining integrity is *cognitive*⁴²³. Adopting this definition would commit Cognitive Science to an object of study to automatically include individual bacteria, cellular collectives, immune systems, ecosystems, markets, governance assemblies, and a substantial portion of what is currently studied under separate disciplinary headings. It would dissolve the biocentric bias

⁴²² Section 1.3.4

⁴²³ With "integrity" meaning what Section 1.3.1 made it mean, "actively" meaning what Section 2.1 made it mean, and "assembly" meaning what Section 1.2 made it mean

that has constrained the field⁴²⁴ and it would require the field to develop empirical methodologies for cognitive systems that do not have the courtesy of being shaped like brains or even like organisms⁴²⁵ at all. Some coordinates of Cognitive Science work definitionally in nervous-system architectures or in specific representational formats, while I expect these will sit uneasily within the redrawn perimeter⁴²⁶, I anticipate that free-energy reducing active goal seekers will sit comfortably within CogSci's frameworks. The payoff is a discipline whose first principles travel cleanly across substrates, whose theoretical vocabulary is portable across scales without the constant translation friction that currently consumes most of the energy at cross-domain conversations, and whose object of study finally matches what was, I suspect, always the underlying intuition that cognition is what non-equilibrium assemblies do to persist, and that wherever such persistence occurs, Cognitive Science has business being and studying.

6.4.2 For Political Science

Constitutions vary enormously in the minimal number of recursive operations required to construct them. A constitution that presupposes literate populations, functioning courts, prior commercial codes, and centuries of jurisprudential precedent has a vastly higher assembly index than one that relies on personalised authority and oral tradition. *Mapping* the assembly index of existing constitutional texts against the cultural assembly index of the populations they govern would, I suspect, reveal a far stronger predictor of constitutional persistence than the ideological-architectural variables Political Science has historically tracked.

This connects directly to what is probably the most consequential research question the framework opens, being this the cultural assembly index condition on constitutional viability. The repeated failures of nation-building over the last several decades⁴²⁷ are difficult to interpret within the dominant frameworks of Political Science without resorting either to cultural-essentialist explanations that the field rightly resists or to institutional-design explanations that fail to predict the same failure modes recurring across radically different institutional designs. The framework developed here offers a third option, that constitutional code requires cultural-scaffolding of sufficient assembly index to interpret and decompress it⁴²⁸ and that importing a constitution whose assembly index exceeds the receiving culture's assembly index produces the same organ-rejection dynamics analysed at Section 2.5.

It can be argued that the Afghan constitution of 2004 was a perfectly good piece of constitutional code by the standards of the donor culture that drafted it. The cultural-scaffolding required to decompress provisions like "independent judiciary," "due process," and "secular citizenship" into context-appropriate operations was, across much of the country, absent. What returned in 2021 was the cultural-scaffolding configuration whose assembly index actually matched the territory's existing cognitive niche. Iraq's post-2003 trajectory tells a structurally similar story, with the imposed federal constitutional architecture decompressing into sectarian patronage networks because those were the dynamic-drive configurations the receiving culture's cognitive-glue actually maintained. I find this reading

⁴²⁴ Even after externalism and extended cognition formally challenged it, I kept running into biocentric biases during my studies at UniWien

⁴²⁵ Feel free to borrow total correlation, singularity-node detection, assembly-index estimation, and all other frameworks leveraged in this thesis

⁴²⁶ I have already faced them from the moment I announced my thesis aim, in plenty of conversations

⁴²⁷ Iraq from 2003, Afghanistan from 2001 to 2021, Libya after 2011, the long arc of Haitian state failure, the post-Yugoslav transitions in different directions across populations whose pre-conflict cultural-scaffolding configurations differed in measurable ways...

⁴²⁸ Sections 3.2.2, 4.1.2

uncomfortable, because it implies that a great deal of well-intentioned international policy has been operating on a structurally flawed model of how constitutional code instantiates. I also find it more empirically faithful than the alternatives, and I think Political Science would benefit from taking it seriously enough to give it a try.

The co-determination of system persistence by hard protocols and dynamic drives is the second research line I would like to see pursued. The thesis argues⁴²⁹ that constitutional code is not self-executing, that its phenotypic expression depends on the dynamic drives of the humans who interpret and enact it, while those drives are simultaneously shaped by the protocols they interpret. Where the framework predicts misalignment between protocol specification and frontline interpretation⁴³⁰ should be predictably non-compliant or selectively enforced. The Portuguese examples I gave for abortion and drug consumption prior to legal reform are anecdotal, but the underlying mechanic should be visible in any jurisdiction where the gap between codified and enforced law is measurable.

The third research line concerns governance externalities as cognitive blind spots. The framework treats variable sensitivity as a structural property of the governance system's cognitive light cone⁴³¹ with externalities as the predictable consequence of variables the system cannot render. Auditing existing governance systems for the variables their architecture is structurally blind to would shift the analytical question from "what is going wrong?" to "what cannot this system see?" I suspect the second framing is more useful, because it points directly at the morphological intervention required (build an organ that renders the missing variable) and skips past the symptomatic interventions that occupy most current policy debate.

Finally, I want to underline a research line that the framework opens, the embodied costs of governance trajectories on the human substrate. Section 4.3 traced the dynamics by which higher-level Hard Protocols project downward as conditioning on lower-level distributions, with the pathologies of sleep deprivation, deaths of despair, and the rest as consequences. Treating these as object of governance analysis, rather than as object of public-health analysis or sociology of medicine, would require Political Science to accept that the embodied condition of the population is not an externality to governance but a phenotype of it.

6.4.3 For Governance Cognitive System science

The most important thing this framework enables is the existence of the field itself. Once governance is recognised as a class of cognitive system continuous with the rest of nature, the research questions that organise the field become structurally similar to those across medicine or neuroscience to computer science, such as "what are the system's sensory channels", "what variables does it render", "how does its generative model update", "what are its characteristic biases", "what failure modes does its architecture predispose toward", "what does its assembly history reveal about why it is built the way it is"...?

Several specific research lines follow. Such as variable-rendering audits across existing governance systems to map what each system can and cannot see. A comparative atlas of governance cognitive light cones, for example, what Sweden renders that the United States does not, what Singapore renders that Brazil does not, what the European Union renders that any of its member states alone do not... I expect such an atlas to reveal blind

⁴²⁹ At Section 4.3

⁴³⁰ High-bandwidth provisions in low-bandwidth cultures, low-bandwidth provisions in cultures whose dynamic drives demand high contextual flexibility

⁴³¹ Section 4.2

spots clustered in characteristic ways, with long-horizon environmental variables under-rendered almost everywhere, cross-generational equity variables under-rendered almost everywhere, multi-scale individual-wellbeing variables under-rendered in proportion to the population over which the system is averaging...

Constitutional architecture as cognitive design is a second line I am excited about. Section 6.2 sketched what designed feedback integration, memory consolidation, attention and metacognition and cross-Markov-blanket model coordination might look like at the governance scale. Each of these is a research programme of its own. I would like to see empirical work on whether legal systems with embedded sunset clauses outperform those without, on whatever metric the legislation was meant to advance. I would like to see comparative analysis of jurisdictions that have implemented explicit feedback integration mechanisms⁴³² against those that have not. I would like to see what happens when a jurisdiction explicitly designs a metacognitive layer, an organ whose function is to model the performance of other organs against long-horizon goals, and whether that organ resists capture by the same short-horizon dynamics that have historically defeated central planning offices.

I also believe *Valley-of-sacrifice traversals*⁴³³ deserves dedicated study. The empirical question of when governance systems successfully cross trajectories whose intermediate readings degrade against their primary sensors and what architectural features make such crossings possible. Which has direct application to climate policy, to long-horizon infrastructure investment, to demographic transitions, to fiscal consolidation, and to most other governance problems that anyone genuinely cares about. The Portuguese Troika example I gave is a single data point, with a complicated outcome, in a research space where dozens or hundreds of well-documented cases could exist. Building a comparative taxonomy of attempted valley traversals, with what worked, what failed, what architectural features correlate with success, would be a substantial contribution to a field that does not yet exist.

The new field will also benefit from studying its own *pathology catalogue*. The thesis hinted at this with the parallels between cellular autoimmunity and bureaucratic compression faults⁴³⁴. A systematic typology of governance-level pathologies, that is, the failure modes a cognitive system of this class is structurally vulnerable to, would be useful both for diagnosis and for design. Some pathologies, like blind-spot externalities, are already partially understood. Others, like the institutional analogues of agnosia (where a system's organs render a variable but cannot integrate it into action, climate science under most twenty-first-century governments is the canonical contemporary example) may not. The framework predicts that pathologies should cluster in characteristic ways across governance architectures, with centralised architectures vulnerable to certain failures (capture, single-point feedback collapse) and distributed architectures vulnerable to others (coordination failure, valley-traversal failure). Mapping that vulnerability landscape would be substantive and interesting work.

Without being exhaustive, a final thought goes to path-dependent assemblies whose viability depends on the bottlenecks crossed in their construction⁴³⁵. Tracing the assembly history of existing constitutional codes, not as legal history in the conventional sense, but as assembly-graph reconstruction of the constraint configurations that successively crystallised,

⁴³² The better-regulation agendas in several European jurisdictions point in this direction, however imperfectly

⁴³³ Section 6.1.4

⁴³⁴ Section 2.5

⁴³⁵ Section 4.1

would reveal which features of any given constitution are load-bearing assembly bottlenecks versus which are downstream artefacts whose modification would not destabilise the overall assembly.

6.4.4 For the AI Era

Plenty of research opportunities streaming from this thesis can directly be addressed by the construction of virtual societies as constitutional laboratories. Large language models can be instantiated as agents with calibrated dynamic drives, cultural-scaffolding priors, hierarchies of values and cognitive light cones of specifiable width. The technology is already adequate to populate a simulated polity with thousands of such agents, expose them to a hard-protocol architecture⁴³⁶ and observe the dynamics that emerge. It enables to run policy interventions *in silico*, observe their downstream effects across multiple coupled timescales and iterate on the constitutional design without imposing the costs of iteration on actual populations. The framework provides the conceptual scaffolding to interpret what is observed, with bowtie compression-decompression cycles, total-correlation diagnostics, singularity-node detection, and valley-of-sacrifice geometries all becoming available as analytical tools applied to the simulated dynamics.

The framework implies that constitutional importation across mismatched cultural assembly pathways should fail in characteristic ways. A virtual society populated by agents with culturally-calibrated dynamic drives exposed to a constitutional design originating in a different cultural pathway should exhibit, in simulation, the same failure modes observed in the historical record. If the simulation reproduces the historical pattern, we have a tool for evaluating proposed constitutional interventions before they are imposed. If the simulation reproduces only some of the pattern, we have learned something specific about what features the framework still does not capture.

A closely related research line concerns the measurement of cultural-scaffolding penetration thresholds. The thesis argues that constitutional viability depends on the cultural-scaffolding penetration across the population the system attempts to govern⁴³⁷. How many people need to share a given cultural-scaffolding for a Governance Cognitive System dependent on that scaffolding to remain viable? The LLM-agent simulator is a natural instrument to answer this, to instantiate a virtual polity of N agents, set the cultural-scaffolding penetration parameter at some fraction p of the population, with the remaining running an alternative scaffolding configuration (or no shared scaffolding at all, or some distribution across several). Expose the polity to a fixed constitutional architecture whose Hard Protocols presuppose the dominant scaffolding, and observe whether the system's Markov blanket holds, that is, whether the constitutional code continues to decompress into context-appropriate operations, whether the higher-level cognitive light cone continues to render the variables the system depends on for its own integrity, basically if it holds in integrity or not. One can experiment sweeping p across a range and see at which shared cultural scaffoldings do Constitutional systems become or remain viable. The framework predicts a phase transition somewhere in the interior and the singularity-node apparatus predicts that the transition will be detectable as a non-analyticity in the joint distribution rather than as a smooth degradation. The location of that transition, and how it varies with the assembly index of the constitutional code, the heterogeneity of the receiving polity, and the architecture of the protocols themselves, is a substantial research programme

⁴³⁶ A candidate constitution, a regulatory regime, an institutional design

⁴³⁷ Sections 3.2.4, 4.1.2

with immediate practical stakes. I suspect that the threshold will turn out to depend less on raw penetration than on the distribution of penetration across functional roles, that a polity in which the scaffolding is concentrated in the offices that interpret and enforce the code may remain viable at population-level penetrations that would otherwise predict collapse, while a polity in which the scaffolding is widely but shallowly distributed may fail at penetrations that look superficially safe.

Calibrating LLM agents to specific cultural assembly indexes is itself a research problem worth pursuing on its own merits. Current LLMs are trained on corpora overwhelmingly skewed toward high-assembly-index cultural-scaffolding (literate, abstract, individualist, late-modern, western aligned...). Have them attempt to model the dynamic drives of a Kpelle farmer or a Pashtun elder poorly. Building agent populations that accurately reflect the cultural-scaffolding diversity of the world is partly a corpus problem, partly a methodological problem, and partly an ethical one. Then one can also attempt different alternative Constitutional assemblies across incremental indexes compatible with those cultural indices to explore alternative governance types.

Another research line follows from the dynamic-drive design implications of Section 6.3. If we can instantiate agent populations with calibrated drives, we can also instantiate the deliberative venues that surface those drives, the value-hierarchy ballots, the time-horizon-indexed preference solicitations, the cognitive-style-stratified panels... We can pilot these in simulation, observe how well the resulting preference structures actually steer system trajectories, and iterate on the architecture before deploying it. This is, in effect, building the prototype of a democratic infrastructure that does not yet exist anywhere in the world. I think such an infrastructure will eventually be necessary regardless, because AI-augmented governance without explicit preference surfacing produces, by structural necessity, the kind of opaque optimisation against under-specified objectives that everyone in AI safety has been worrying about. Surfacing the objectives at the polity level is, in this view, the same problem as alignment.

I want to close with what I take to be the most important question the framework enables, and the one I most want future researchers to carry forward. As AI capabilities increase, the cognitive light cone of governance cognitive systems will expand dramatically. Systems that previously could not render certain variables will, with sufficient AI integration, become able to render them. The question I most want to see pursued is whether the dynamic drives of the human substrate can keep up. The thesis argues that hard protocols and dynamic drives co-determine system viability⁴³⁸ and that misalignment in either direction produces pathology. If AI-driven hard-protocol generation outpaces the cultural-scaffolding evolution required to interpret and embody those protocols, the prediction the framework makes is one of accelerating organ-rejection dynamics, that is, the risk of more Robodebts, more Windrushes, more situations in which the codified system runs trajectories that the embodied substrate cannot tolerate or perhaps even represent. Nobody knows how to slow down once AI governance is enabled, so the first fundamental step is to be able to see it.

⁴³⁸ Section 2.6

6.5 Conclusion

The boundaries that definitions impose on what they attempt to capture are themselves subject to the assembly pathway that led to their specific angle of inquiry. Cognition, governance, system, agent, self... each of these concepts is held together more by how we come to interpret them than by how they describe the world. To actually perform the thesis above, I knew I had to drill down into the phenomena that cross conceptual lines, with disregard for settled academic definitions and a willingness to lean on a well-established core of principles I could leverage functionally.

This was the guiding thought that led me to the pursuit of a common analytical substrate capable of bridging Cognitive Science and Political Science from first principles, in response to the diagnosis that each discipline, on its own, either lacked or contaminated the vocabulary needed to address what merges them. The contribution this thesis sets out can be restated in five interlocking claims.

First, that information⁴³⁹ provides a reductionist substrate that both traverses and defines levels and substrates, with positive total correlation serving as the diagnostic of systemhood and recursive coarse-graining as the grammar of level nesting. Second, that the temporal accumulation of constraints across path-dependent assembly space and their consolidation (in singularity-nodes) together provide a non-teleological account of how new levels of organisation form and persist, with selection reread as the statistical reweighting of probability mass over assembly paths in favour of those whose terminal configurations propagate. Third, that once non-equilibrium thresholds are crossed, Cognition becomes definable as *what any assembly does to actively maintain its integrity* through prediction-error minimisation. Fourth, that Constitutions are recursive assembly memory⁴⁴⁰ and that Governance Cognitive Systems are higher-order Markov blankets whose generative models, variable sensitivities, attentional regimes and pathologies are tractable objects of empirical study. And finally that hard-protocols, dynamic-drives and assembly pathways/indexes operate as co-determinants of system viability across a measurable parameter space. Externalities, institutional pathologies, valley-of-sacrifice traversals, the embodied costs of governance trajectories on their human substrate, and the conditions under which constitutional code remains viable on a given cultural-scaffolding assembly index all become tractable from this single vantage point.

I am the first to acknowledge that the probability-space machinery is demanding to deploy and will require its own line of research to become functionally tractable, and that the singularity-node criteria are easier to state than to estimate and equally require further work. I also readily advance that the political science I have engaged with is the one I happened to predominantly study during my Law degree. None of this troubles me critically, as I expect future iteration on the framework to refine what I consider to already have utility, being happy once the framework has proved daring enough to be iterated upon, since I remain with the impression of having pointed in the right direction.

What I will say in closing is that neither of the disciplines I trained in were accidents. Application forms show how from day one I intended to use my training towards this aim. I arrived at the conclusion that Law treated the state as if it had no body and Cognitive Science treated the body as if it had, and wasn't itself the embodiment of, a state. If through this thesis the only thing I imprinted into the reader is the equipment to contest that state of

⁴³⁹ Defined as the measurable departure of a joint distribution from its maximum-entropy reference, on whatever probability space the level of inquiry requires

⁴⁴⁰ Code encoding both its first-order constraints and the procedures for its own iteration

affairs, then it is enough for me to regard all this effort as successful. Personally, what I wanted from this thesis was a place to feel as having been a part of the things I find most consequential about the world. I hope to spend a considerable portion of my professional, academic and personal life doing precisely that.

Glossary

In order of appearance

Reductionist Substrate (Preamble, 1.3.4) — The elemental building block from which a given reductionist approach derives its analysis and explanations. In this thesis, the substrate committed to is *information* as the measurable departure of a joint distribution from its maximum-entropy reference, on whatever probability space the level of inquiry requires.

Cartesian Architecture / Dualism (Preamble, 5.2.2) — Inherited assumption that institutions, constitutions and the populations they govern are categorically distinct entities, separable for analytical purposes and ultimately differentiable in their ontology. Descended from Descartes' separation of *res extensa* from *res cogitans*, this dualist architecture organises much of post-war legal and political thought (Kelsen, Hart, Rawls, Habermas, Luhmann) and is rejected in this thesis in favour of a monist account in which constitutional code and human substrate stand in constitutive relation.

GST (General Systems Theory) (1.1) — Framework emphasising the structural organisation of reality in terms of interconnected and hierarchical nested systems rather than isolated parts; the precursor to the informational-systemic approach developed throughout the thesis.

Probability Space (1.1.1) — Triple of sample space, collection of events, and probability measure. The formal object on which every information-theoretic quantity in this thesis is defined.

Sample Space (1.1.1) — The set of all microstates available to a system under analysis, fixed by the level of inquiry.

Microstate / Alphabet of States (1.1.3) — A fully specified, fine-grained configuration of a system or its parts. The alphabet is the set of all such configurations a system can occupy.

Information (1.1.2) — The measurable departure of a joint distribution from its maximum-entropy reference, capturing both temporal acquisition of knowledge and spatial instantiation of order. Carried by three complementary registers: negentropy, mutual information, and total correlation.

Negentropy (1.1.2) — The constraint content of a single distribution; its measurable departure from the uniform (maximum-entropy) reference.

Mutual Information (1.1.2) — Quantity capturing how much knowing the value of one random variable reduces uncertainty about another; used here for cross-level coupling.

Total Correlation (1.1.2, 1.1.3) — Measure of coupling among the components of a system at one level; the diagnostic of systemhood adopted in place of variance-based formulations.

Weiss's System (1.1) — Historical/intuitive precursor to the informational definition of system, holding that a given complex is a system when the variance of the whole is less than the sum of variances of its constituents. Superseded in this thesis by positive total correlation (1.1.3).

System (1.1.3) — A multi-component level whose total correlation is positive. Its joint distribution is concentrated on a strictly smaller alphabet than the Cartesian product of the component alphabets would allow.

Coarse-graining (1.1.4) — Function partitioning a fine-grained microstate alphabet into equivalence classes, replacing a distribution over a larger alphabet with one over a smaller alphabet. The formal grammar of *level nesting*.

System Level Integrity (1.3.1, 2.1) — Maintenance of structural coherence of a system under the perturbations it actually encounters; the criterion against which assembly paths are reweighted.

Assembly Theory / AT (1.2) — Framework reconceptualising objects as entities defined by their formation histories in an assembly space, with their identity being inseparable from path-dependent construction.

Assembly History (1.2) — The path-dependent sequence of constraint applications by which an object came to exist. Functions as systemic memory enabling reuse of stabilised solutions.

Assembly Index (1.2) — The minimal number of recursive compositional operations required to construct an object from elementary building blocks.

Copy Number (1.2) — The number of identical instances of a given object; together with assembly index, signals selection acting on historically realised construction paths.

Assembly Bottleneck (1.2.1) — A substructure in assembly space through which paths to many distinct higher-index objects must pass. Exhibits convergent constraint satisfaction, reusability amplification, and informational compression.

Constraint Family (1.3.2) — One-parameter family of distributions over the same components, indexed by a parameter governing the constraints.

Phase Transition (1.3.2) — Critical parameter value at which the free energy of a constraint family becomes non-analytic. One of the two formalisations of singularity-nodes.

Information Bottleneck Bifurcation (1.3.2) — Critical value of the compression-relevance trade-off at which the optimal representation undergoes qualitative change. The parameter-free analogue of phase transitions.

Singularity-node (1.3.3) — Configuration where a constraint family ceases to be analytic in its parameters (or where the IB Lagrangian bifurcates). The site of level-formation.

Path-reweighting / Selection (1.3.4) — Statistical reweighting of probability mass over assembly paths in favour of those whose terminal configurations persist and propagate.

Downward Causation (Preamble, 1.3.4) — Influence of higher levels on lower-level configurations, formalised as conditioning of $P(X)$ on higher-level state Y to yield $P(X|Y)$.

Non-equilibrium Assembly (2.1) — Assembly configuration that, having crossed the bottleneck from passive to active constraints, must actively maintain its integrity through ongoing energy expenditure rather than passive constraint satisfaction.

Free Energy (1.3.2, 2.1) — Quantity summarising the trade-off between energy minimisation and entropy maximisation under constraints; minimised statically to locate singularity-nodes and dynamically by active-inference agents.

Active Inference (2.1) — Framework in which non-equilibrium systems persist by minimising prediction error through coordinated perception-action cycles.

Generative Model (2.1) — Internal representation of how the world works that enables a system to predict consequences and bias action toward integrity-maintaining outcomes.

Preferences / Priors (2.1) — Non-uniform prior distributions encoding which states a system systematically biases toward; the origin of cognitive goals.

Valence (2.1, 4.2) — Systematic bias toward or away from particular stimuli/conditions; statistical classification of distance from preferred coordinates.

Multiscale Competency Architecture (2.2) — Nested arrangement in which each level retains autonomous problem-solving capacities while contributing to higher-level integrity (term from Levin).

Task Environment (2.4) — Framework (Newell & Simon) defining the environment of a cognitive system relative to its tasks and goals. Anchored in 3D spatial navigation and object manipulation; superseded in this thesis by *Problem Space*, which generalises the concept across non-sensorimotor problem dimensions and across scales.

Affordance (2.4) — Concept from ecological psychology (Gibson) denoting the action possibilities an environment offers to an agent. Relational rather than intrinsic, dependent jointly on the environment's properties, the agent's capabilities, and its goal-state. Critiqued here as too narrowly anchored to immediate sensorimotor navigation to capture the multidimensional problem-spaces cognitive systems actually inhabit.

Cognitive Light Cone (2.4) — Spatio-temporal and variable scope within which a system can sense, model, predict, and act (term from Levin).

Markov Blanket (2.4) — Statistical partition demarcating internal states from external environment via blanket states filtering exchanges; defines systemhood informationally rather than morphologically.

Problem Space (2.4) — Multidimensional possibility landscape (metabolic, transcriptional, governance, etc.) within which a system navigates, broader than 3D task-environments.

Bureaucracy (2.5) — Functionally, a scale-free class of informational solutions to coordination problems in multi-agent systems, implemented through bowtie compression-decompression cycles.

Bowtie Architecture (2.5) — Information-processing pattern in which multidimensional inputs are compressed into low-dimensional bottleneck representations, then decompressed contextually into outputs.

Bowtie Bottleneck (2.5) — The compressed central node within a bowtie architecture — the low-dimensional representation through which multidimensional inputs are channelled before being decompressed contextually. Examples include p53 in cellular stress response, corticotropin-releasing hormone in the HPA axis, planarian bioelectric pattern memory, and constitutional documents in governance assemblies.

Hard Protocols (2.6) — Mechanistic constraints, architectural structures, and regulatory thresholds delimiting the joint possibility space of multi-agent systems.

Dynamic Drives (2.6) — Adaptive interpretation processes through which agents navigate constrained spaces while pursuing their own free-energy minimisation.

Functionalism (2.7) — Methodological commitment to defining and classifying entities by the *role* they play in constraint propagation, uncertainty reduction, and integrity maintenance, rather than by the substrate through which those roles are instantiated. Generalises the parity principle and frees terminology from substrate-bound legacy across cellular, organismal, cultural and governance scales.

Cognitive Glue (2.7, 3.2.4) — Informational medium binding distributed agents into integrated cognitive systems by establishing collective-level conditional dependencies and a higher-level Markov blanket.

Scale-free Cognition (2.1, 2.7) — The thesis that cognition, as constraint propagation toward integrity in non-equilibrium systems, is substrate-independent and instantiates the same operational principles at any level of organisation.

Cognitive Scaffolding (3.1) — External physical, cognitive, or social structures that augment individual cognition by offloading or distributing computational load.

Cognitive Uploading (3.2.4) — Generational process of actively tailoring shared environments to support cognitive operations at scale.

Culture (3.2.4) — Shared pool of collectively maintained external cognitive scaffoldings pruning surprise in individual and collective predictive processes.

Abstraction Bottleneck (3.2.2) — Cultural-scaffolding assembly index required for generalised, context-independent symbolic manipulation.

Baldwin Effect (3.2.4, 4.3) — Mechanism by which learned/cultural behaviours channel subsequent genetic evolution toward traits supporting them.

Constitution (4.1) — Self-referential assembly encoding both first-order constraints and procedures for modifying those constraints — thus recursive assembly memory (achieved at governance assemblies).

Governance Cognitive System (4.2) — Higher-order assembly whose collective Markov blanket emerges from human agents bound by cultural scaffolding and constitutional code. Its cognitive operations — variable sensitivity, generative modelling, integrity drives, attention, memory — are irreducible to those of its constituent humans, and its trajectory through problem-space is mediated by morphological self-modification at the constitutional layer.

Variable Sensitivity (4.2) — Range of variables a governance system can render and act upon; determined by its morphology (organs, registries, instruments).

Externality (4.2) — Costs or benefits a governance system fails to render through its variable sensitivities, producing systematic biases that manifest at the system level but are invisible from any individual agent's perspective. Diagnosable as cognitive blind spots arising from compression artifacts in the system's bowtie architecture or from absent organs that would otherwise render the relevant variables.

Red Queen Effect (4.3) — Coevolutionary dynamic in which lower-level agents face escalating adaptation pressure to remain compatible with higher-level system fit.

Valley of Sacrifice (6.1.4) — Topological feature where local and global gradients in problem-space point in opposite directions, traversable only by systems whose cognitive light cone encompasses the far rim.

References

- Bacon, F. (2000). *The new organon* (L. Jardine & M. Silverthorne, Eds.). Cambridge University Press. (Original work published 1620)
- Baluška, F., & Levin, M. (2016). On having no head: Cognition throughout biological systems. *Frontiers in Psychology*, 7, 902. <https://doi.org/10.3389/fpsyg.2016.00902>
- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, 133(1), 1–24. <https://doi.org/10.1037/0033-2909.133.1.1>
- Bar-Yam, Y. (2004). A mathematical theory of strong emergence using multiscale variety. *Complexity*, 9, 15–24. <https://doi.org/10.1002/cplx.20029>
- Barsalou, L. W. (2008). Cognitive and neural contributions to understanding the conceptual system. *Current Directions in Psychological Science*, 17(2), 91–95. <https://doi.org/10.1111/j.1467-8721.2008.00555.x>
- Beer, S. (1981). *Brain of the firm* (2nd ed.). John Wiley & Sons.
- Beetz, M. J., Kraus, C., & el Jundi, B. (2023). Neural representation of goal direction in the monarch butterfly brain. *Nature Communications*, 14, Article 5859. <https://doi.org/10.1038/s41467-023-41526-w>
- Bellinger, D. C., Devleesschauwer, B., O'Leary, K., & Gibb, H. J. (2016). Country-specific estimates of the incidence of intellectual disability associated with prenatal exposure to methylmercury. *Environmental Research*, 147, 159–163. <https://doi.org/10.1016/j.envres.2015.10.006>
- Berg-Schlosser, D., Badie, B., & Morlino, L. (Eds.). (2020). *The SAGE handbook of political science* (Vol. 1). SAGE Publications.
- Bizzi, E., Mussa-Ivaldi, F. A., & Giszter, S. (1991). Computations underlying the execution of movement: A biological perspective. *Science*, 253(5017), 287–291. <https://doi.org/10.1126/science.1857964>
- Blackiston, D., Kriegman, S., Garnier, S., Bongard, J., & Levin, M. (2021). A cellular platform for the development of synthetic living machines. *Science Robotics*, 6(52), eabf1571. <https://doi.org/10.1126/scirobotics.abf1571>
- Blackiston, D. J., & Levin, M. (2013). Ectopic eyes outside the head in *Xenopus* tadpoles provide sensory data for light-mediated learning. *Journal of Experimental Biology*, 216(6), 1031–1040. <https://doi.org/10.1242/jeb.074963>

- Blackiston, D. J., McLaughlin, K. A., & Levin, M. (2009). Bioelectric controls of cell proliferation: Ion channels, membrane voltage, and the cell cycle. *Cell Cycle*, 8(21), 3519–3528. <https://doi.org/10.4161/cc.8.21.9888>
- Blankenship, R. E. (2010). Early evolution of photosynthesis. *Plant Physiology*, 154(2), 434–438. <https://doi.org/10.1104/pp.110.161687>
- Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, 36(1), 75–111. <http://www.jstor.org/stable/2564952>
- Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511804441>
- Brown, G. C. (1992). Control of respiration and ATP synthesis in mammalian mitochondria and cells. *Biochemical Journal*, 284(Pt 1), 1–13. <https://doi.org/10.1042/bj2840001>
- Brown, S. M. (1963). [Review of *The concept of law*, by H. L. A. Hart]. *The Philosophical Review*, 72(2), 250–253. <https://doi.org/10.2307/2183110>
- Cacioppo, J. T., & Cacioppo, S. (2018). The growing problem of loneliness. *The Lancet*, 391(10119), 426.
- Campbell, D. T. (1974). Downward causation in hierarchically organized biological systems. In F. J. Ayala & T. Dobzhansky (Eds.), *Studies in the philosophy of biology: Reduction and related problems* (pp. 179–186). Macmillan. https://doi.org/10.1007/978-1-349-01892-5_11
- Case, A., & Deaton, A. (2020). *Deaths of despair and the future of capitalism*. Princeton University Press.
- Castellanos, I., Barbosa, P., Zuria, I., Tammaru, T., & Christman, M. C. (2011). Contact with caterpillar hairs triggers predator-specific defensive responses. *Behavioral Ecology*, 22(5), 1020–1025. <https://doi.org/10.1093/beheco/arr085>
- Chen, Y., Xu, J., & Chen, Y. (2021). Regulation of neurotransmitters by the gut microbiota and effects on cognition in neurological disorders. *Nutrients*, 13(6), Article 2099. <https://doi.org/10.3390/nu13062099>
- Chialvo, D. R. (2010). Emergent complex neural dynamics. *Nature Physics*, 6(10), 744–750. <https://doi.org/10.1038/nphys1803>
- Clark, A. (1997). *Being there: Putting brain, body, and world together again*. MIT Press.
- Clark, A. (2012). Embodied, embedded, and extended cognition. In K. Frankish & W. Ramsey (Eds.), *The Cambridge handbook of cognitive science*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139033916>

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1093/analys/58.1.7>

Clark, A., & Toribio, J. (Eds.). (1998). *Artificial intelligence and cognitive science: Volume 4. Language and meaning in cognitive science*. Garland Publishing. (Reprint 2011 by Routledge)

Conrad, P., & Bergey, M. R. (2014). The impending globalization of ADHD: Notes on the expansion and growth of a medicalized disorder. *Social Science & Medicine*, 122, 31–43.

Constant, A., Clark, A., & Friston, K. J. (2021a). Representation wars: Enacting an armistice through active inference. *Frontiers in Psychology*, 11, Article 598733. <https://doi.org/10.3389/fpsyg.2020.598733>

Constant, A., Clark, A., Kirchhoff, M., & Friston, K. J. (2022). Extended active inference: Constructing predictive cognition beyond skulls. *Mind & Language*, 37(3), 373–394. <https://doi.org/10.1111/mila.12330>

Constant, A., Tschantz, A. D. D., Millidge, B., Criado-Boado, F., Martinez, L. M., Müller, J., & Clark, A. (2021b). The acquisition of culturally patterned attention styles under active inference. *Frontiers in Neurobotics*, 15, Article 729665. <https://doi.org/10.3389/fnbot.2021.729665>

Constituição da República Portuguesa. (1976). *Diário da República*. <https://diariodarepublica.pt/dr/legislacao-consolidada/decreto-aprovacao-constituicao/1976-34520775-48383075>

Cottingham, J. (Ed.). (1992). *The Cambridge companion to Descartes*. Cambridge University Press.

Couto, A., Wainwright, J. B., Morris, B. J., & Montgomery, S. H. (2020). Linking ecological specialisation to adaptations in butterfly brains and sensory systems. *Current Opinion in Insect Science*, 42, 55–60. <https://doi.org/10.1016/j.cois.2020.09.002>

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley-Interscience.

Curiel, E. (2019). Light cones and causal structure. In *Singularities and black holes* (Supplement). *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/spacetime-singularities/lightcone.html>

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Prentice-Hall.

Dadvand, P., Nieuwenhuijsen, M. J., Esnaola, M., Forns, J., Basagaña, X., Alvarez-Pedrerol, M., Rivas, I., López-Vicente, M., De Castro Pascual, M., Su, J., Jerrett, M., Querol, X., & Sunyer, J. (2015). Green spaces and cognitive development in primary schoolchildren. *Proceedings of the National Academy of Sciences*, 112(26), 7937–7942.

Daniel, K. L., McConnell, M., Schuchardt, A., & Pepper, M. E. (2022). Challenges facing interdisciplinary researchers: Findings from a professional development workshop. *PLOS ONE*, 17(4), Article e0267234. <https://doi.org/10.1371/journal.pone.0267234>

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511524059>

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17), 6889–6892. <https://doi.org/10.1073/pnas.1018033108>

Davies, P., & Gregersen, N. H. (Eds.). (2010). *Information and the nature of reality: From physics to metaphysics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511778759>

Day, C. (2024, May 16). How water flows inside a sea sponge. *Physics*, 17, Article 81. <https://physics.aps.org/articles/v17/81>

Dennett, D. C. (1991). *Consciousness explained*. Little, Brown and Company.

Deutsch, K. W. (1966). *The nerves of government: Models of political communication and control*. Free Press.

Devaraj, S., Hemarajata, P., & Versalovic, J. (2013). The human gut microbiome and body metabolism: Implications for obesity and diabetes. *Clinical Chemistry*, 59(4), 617–628. <https://doi.org/10.1373/clinchem.2012.187617>

Devarajan, P., & Chen, Z. (2013). Autoimmune effector memory T cells: The bad and the good. *Immunological Research*, 57, 12–22. <https://doi.org/10.1007/s12026-013-8448-1>

Dias, D. A., Robalo Marques, C., & Richmond, C. (2014). *Misallocation and productivity in the lead up to the Eurozone crisis* (Working Paper No. w201411). Banco de Portugal. <https://www.bportugal.pt/paper/misallocation-and-productivity-lead-eurozone-crisis>

Driscoll, C. A., Macdonald, D. W., & O'Brien, S. J. (2009). From wild animals to domestic pets, an evolutionary view of domestication. *Proceedings of the National Academy of Sciences*, 106(Supplement 1), 9971–9978. <https://doi.org/10.1073/pnas.0901586106>

Durant, F., Morokuma, J., Fields, C., Williams, K., Adams, D. S., & Levin, M. (2017). Long-term, stochastic editing of regenerative anatomy via targeting endogenous bioelectric gradients. *Biophysical Journal*, 112(10), 2231–2243. <https://doi.org/10.1016/j.bpj.2017.04.011>

Ebrahimkhani, M. R., & Levin, M. (2021). Synthetic living machines: A new window on life. *iScience*, 24(5), 102505. <https://doi.org/10.1016/j.isci.2021.102505>

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, 91(4), 451–471. <https://doi.org/10.1348/000712600161934>

Endo, H., Tsuneto, K., Mang, D., Zhang, W., Yamagishi, T., Ito, K., Nagata, S., & Sato, R. (2024). Molecular basis of host plant recognition by silkworm larvae. *Journal of Insect Physiology*, 154, 104628. <https://doi.org/10.1016/j.jinsphys.2024.104628>

Eurostat. (n.d.). *Information on data*. European Commission. Retrieved July 22, 2025, from <https://ec.europa.eu/eurostat/web/hicp/information-data>

Faqeeh, A., Osat, S., Radicchi, F., & Gleeson, J. P. (2019). Emergence of power laws in noncritical neuronal systems. *Physical Review E*, 100(1-1), 010401. <https://doi.org/10.1103/PhysRevE.100.010401>

Farnam Street. (n.d.). *This is water by David Foster Wallace (full transcript and audio)*. Retrieved August 9, 2024, from <https://fs.blog/david-foster-wallace-this-is-water/>

Fernandez Velasco, P. (2023). Entropy, prediction and the cultural ecosystem of human cognition: A challenge to the union of distributed cognition and predictive processing. *Synthese*, 201, 94. <https://doi.org/10.1007/s11229-023-04089-3>

Fields, C., Glazebrook, J. F., & Levin, M. (2021). Minimal physicalism as a scale-free substrate for cognition and consciousness. *Neuroscience of Consciousness*, 7(2), Article niab013. <https://doi.org/10.1093/nc/niab013>

Figueiredo, A., & Silva, J. (2014). The abortion referendum in Portugal: The influence of psychosocial variables in the voting intentions and behaviour of Portuguese youth. *Contention: The Multidisciplinary Journal of Social Protest*, 2(1), 5–24. <https://doi.org/10.3167/cont.2014.020102>

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Franck, C., Grandi, S. M., & Eisenberg, M. J. (2013). Agricultural subsidies and the American obesity epidemic. *American Journal of Preventive Medicine*, 45(3), 327–333. <https://doi.org/10.1016/j.amepre.2013.04.010>

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, 108(3), 819–824. <https://doi.org/10.1016/j.cognition.2008.04.007>

Fridman, L. (2024, June 13). *Sara Walker: Physics of life, time, complexity, and aliens* [Video; Lex Fridman Podcast #433]. YouTube. <https://www.youtube.com/watch?v=wwhTfyX9J34>

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>

Fukuyama, F. (2011). *The origins of political order: From prehuman times to the French Revolution*. Farrar, Straus and Giroux.

Galofré-Vilà, G. (2023). Spoils of war: The political legacy of the German hyperinflation. *Explorations in Economic History*, 88, 101479. <https://doi.org/10.1016/j.eeh.2022.101479>

Gao, F., Guo, J., & Shen, Y. (2024). Advances from chlorophyll biosynthesis to photosynthetic adaptation, evolution and signaling. *Plant Stress*, 12, Article 100470. <https://doi.org/10.1016/j.stress.2024.100470>

Gaspar, C. M. (1999). Habermas' theory of communicative action [Review of *The theory of communicative action*, by J. Habermas & T. McCarthy]. *Philippine Studies*, 47(3), 407–425. <http://www.jstor.org/stable/42634329>

Gay, J., & Cole, M. (1967). *The new mathematics and an old culture: A study of learning among the Kpelle of Liberia*. Holt, Rinehart and Winston.

Geertz, C. (1973). *The interpretation of cultures*. Basic Books.

Genovese, I., Vezzani, B., Danese, A., Modesti, L., Vitto, V. A. M., Corazzi, V., Pelucchi, S., Pinton, P., & Giorgi, C. (2020). Mitochondria as the decision makers for cancer cell fate: From signaling pathways to therapeutic strategies. *Cell Calcium*, 92, 102308. <https://doi.org/10.1016/j.ceca.2020.102308>

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Gogna, R., Shee, K., & Moreno, E. (2015). Cell competition during growth and regeneration. *Annual Review of Genetics*, 49, 697–718. <https://doi.org/10.1146/annurev-genet-112414-055214>

Goldenfeld, N. (1992). *Lectures on phase transitions and the renormalization group*. Addison-Wesley.

Gordeeva, A. V., Labas, Y. A., & Zvyagilskaya, R. A. (2004). Apoptosis in unicellular organisms: Mechanisms and evolution. *Biochemistry (Moscow)*, 69(10), 1055–1066. <https://doi.org/10.1023/B:BIRY.0000046879.54211.ab>

Gumuskaya, G., Srivastava, P., Cooper, B. G., Lesser, H., Semegran, B., Garnier, S., & Levin, M. (2023). Motile living biobots self-construct from adult human somatic progenitor seed cells. *Advanced Science*, 11(4), 2303575. <https://doi.org/10.1002/adv.202303575>

Hafner, M., Stepanek, M., Taylor, J., Troxel, W. M., & van Stolk, C. (2017). *Why sleep matters — the economic costs of insufficient sleep: A cross-country comparative analysis*. RAND Corporation.

Harari, Y. N. (2015). *Sapiens: A brief history of humankind*. Harper.

Harari, Y. N. (2024). *Nexus: A brief history of information networks from the Stone Age to AI*. Random House.

Herman, J. P., Figueiredo, H., Mueller, N. K., Ulrich-Lai, Y., Ostrander, M. M., Choi, D. C., & Cullinan, W. E. (2003). Central mechanisms of stress integration: Hierarchical circuitry controlling hypothalamo–pituitary–adrenocortical responsiveness. *Frontiers in Neuroendocrinology*, 24(3), 151–180. <https://doi.org/10.1016/j.yfrne.2003.07.001>

Hinshaw, S. P., & Scheffler, R. M. (2014). *The ADHD explosion: Myths, medication, money, and today's push for performance*. Oxford University Press.

Hodgson, G. M. (2007). Meanings of methodological individualism. *Journal of Economic Methodology*, 14(2), 211–226. <https://doi.org/10.1080/13501780701394094>

Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, 10(2), 227–237.

Hutchins, E. (1995a). How a cockpit remembers its speeds. *Cognitive Science*, 19(3), 265–288. [https://doi.org/10.1016/0364-0213\(95\)90020-9](https://doi.org/10.1016/0364-0213(95)90020-9)

Hutchins, E. (1995b). *Cognition in the wild*. MIT Press.

Hutchins, E. (2010). Cognitive ecology. *Topics in Cognitive Science*, 2(4), 705–715. <https://doi.org/10.1111/j.1756-8765.2010.01089.x>

Hutchins, E. (2013). The cultural ecosystem of human cognition. *Philosophical Psychology*. <https://doi.org/10.1080/09515089.2013.830548>

Hutchins, E., & Klausen, T. (1996). Distributed cognition in an airline cockpit. In Y. Engeström & D. Middleton (Eds.), *Cognition and communication at work* (pp. 15–34). Cambridge University Press. <https://doi.org/10.1017/CBO9781139174077.002>

Inaba, M., Yamanaka, H., & Kondo, S. (2012). Pigment pattern formation by contact-dependent depolarization. *Science*, 335(6072), 677. <https://doi.org/10.1126/science.1212821>

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4), 620–630. <https://doi.org/10.1103/PhysRev.106.620>

Johnson, P. E. (2006). The politics of attention: How government prioritizes problems. *Perspectives on Politics*, 4(3), 598–599. <https://doi.org/10.1017/S1537592706470365>

Jones, B. D. (2001). *Politics and the architecture of choice: Bounded rationality and governance*. University of Chicago Press.

Jones, R. H. (2013). *Analysis and the fullness of reality: An introduction to reductionism and emergence*. Jackson Square Books.

Kanoho, P., Brannon, J., Prellberg, M., & Baker, E. L. (2017). *Panacea or pipe dream: Does Portugal's drug decriminalization policy translate for Hawaii?* (Report No. 1). Legislative Reference Bureau.

Kelsen, H., & Knight, M. (1967). *Pure theory of law* (1st ed.). University of California Press. <https://doi.org/10.2307/jj.13167921>

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, 15(138), 20170792. <https://doi.org/10.1098/rsif.2017.0792>

Kirkbride, J. B., Anglin, D. M., Colman, I., Dykxhoorn, J., Jones, P. B., Patalay, P., Pitman, A., Soneson, E., Steare, T., Wright, T., & Griffiths, S. L. (2024). The social determinants of mental health and disorder: Evidence, prevention and recommendations. *World Psychiatry*, 23(1), 58–90. <https://doi.org/10.1002/wps.21160>

Kivimäki, M., Jokela, M., Nyberg, S. T., Singh-Manoux, A., Fransson, E. I., Alfredsson, L., et al. (2015). Long working hours and risk of coronary heart disease and stroke: A systematic review and meta-analysis of published and unpublished data for 603,838 individuals. *The Lancet*, 386(10005), 1739–1746.

Klijn, E. H., & Korthagen, I. (2018). Governance and media attention: A research agenda about how media affect (network) governance processes. *Perspectives on Public Management and Governance*, 1(2), 103–113. <https://doi.org/10.1093/ppmgov/gvx004>

Knoblauch, M., Peters, W. S., Bell, K., Ross-Elliott, T. J., & Oparka, K. J. (2018). Sieve-element differentiation and phloem sap contamination. *Current Opinion in Plant Biology*, 43, 43–49. <https://doi.org/10.1016/j.pbi.2017.12.008>

Kool, E. T. (2001). Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annual Review of Biophysics and Biomolecular Structure*, 30(1), 1–22. <https://doi.org/10.1146/annurev.biophys.30.1.1>

Kramarenko, V., Engstrom, L., Verdier, G., Fernandez, G., Oppers, S. E., Hughes, R., McHugh, J., & Coats, W. (2010). *Zimbabwe: Challenges and policy options after hyperinflation* (IMF Working Paper No. 10/3). International Monetary Fund.

Kroupin, I., & Zeng, T. C. (2024, August 13). *From richer to leaner cultures and minds: How depth of standardization as an evolutionary dynamic drives changes in technology, social structure, and cross-cultural cognitive variation*. PsyArXiv. <https://doi.org/10.31234/osf.io/mxfcv>

Kvenvolden, K., Lawless, J., Pering, K., Peterson, E., Flores, J., Ponnampereuma, C., Kaplan, I. R., & Moore, C. (1970). Evidence for extraterrestrial amino-acids and

hydrocarbons in the Murchison meteorite. *Nature*, 228(5275), 923–926. <https://doi.org/10.1038/228923a0>

Lach, D. (2014). Challenges of interdisciplinary research: Reconciling qualitative and quantitative methods for understanding human–landscape systems. *Environmental Management*, 53(1), 88–93. <https://doi.org/10.1007/s00267-013-0115-8>

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3), 183–191. <https://doi.org/10.1147/rd.53.0183>

Lane, D., & Levine, A. (2010). p53 research: The past thirty years and the next thirty years. *Cold Spring Harbor Perspectives in Biology*, 2(12), a000893. <https://doi.org/10.1101/cshperspect.a000893>

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>

Levin, M. (2013). Reprogramming cells and tissue patterning via bioelectrical pathways: Molecular mechanisms and biomedical opportunities. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 5(6), 657–676. <https://doi.org/10.1002/wsbm.1236>

Levin, M. (2014). Molecular bioelectricity: How endogenous voltage potentials control cell behavior and instruct pattern regulation in vivo. *Molecular Biology of the Cell*, 25(24), 3835–3850. <https://doi.org/10.1091/mbc.E13-12-0708>

Levin, M. (2019). The computational boundary of a "self": Developmental bioelectricity drives multicellularity and scale-free cognition. *Frontiers in Psychology*, 10, Article 2688. <https://doi.org/10.3389/fpsyg.2019.02688>

Levin, M. (2023a). Bioelectric networks: The cognitive glue enabling evolutionary scaling from physiology to mind. *Animal Cognition*, 26(6), 1865–1891. <https://doi.org/10.1007/s10071-023-01780-3>

Levin, M. (2023b). Darwin's agential materials: Evolutionary implications of multiscale competency in developmental biology. *Cellular and Molecular Life Sciences*, 80(6), 142. <https://doi.org/10.1007/s00018-023-04790-z>

Levin, M. (2024). Self-improvising memory: A perspective on memories as agential, dynamically reinterpreting cognitive glue. *Entropy*, 26(6), 481. <https://doi.org/10.3390/e26060481>

Levin, M., & Martyniuk, C. J. (2018). The bioelectric code: An ancient computational medium for dynamic control of growth and form. *BioSystems*, 164, 76–93. <https://doi.org/10.1016/j.biosystems.2017.08.009>

- Li, T., & Zhou, X. (2014). The soldiers in societies: Defense, regulation, and evolution. *International Journal of Biological Sciences*, 10(3), 296–308. <https://doi.org/10.7150/ijbs.6847>
- Lindqvist, L. M., Tandoc, K., Topisirovic, I., & Furic, L. (2018). Cross-talk between protein synthesis, energy metabolism and autophagy in cancer. *Current Opinion in Genetics & Development*, 48, 104–111. <https://doi.org/10.1016/j.gde.2017.11.003>
- Linkenauger, S. A., Witt, J. K., Stefanucci, J. K., Bakdash, J. Z., & Proffitt, D. R. (2009). The effects of handedness and reachability on perceived distance. *Journal of Experimental Psychology: Human Perception and Performance*, 35(6), 1649–1660. <https://doi.org/10.1037/a0016875>
- Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The active inference approach to ecological perception: General information dynamics for natural and artificial embodied cognition. *Frontiers in Robotics and AI*, 5, Article 21. <https://doi.org/10.3389/frobt.2018.00021>
- Lobikin, M., Chernet, B., Lobo, D., & Levin, M. (2012). Resting potential, oncogene-induced tumorigenesis, and metastasis: The bioelectric basis of cancer in vivo. *Physical Biology*, 9(6), 065002. <https://doi.org/10.1088/1478-3975/9/6/065002>
- Long, M., Long, E., & Waters, T. (2013). *Suicide among the Mla Bri hunter-gatherers of northern Thailand*. [Manuscript].
- Luhmann, N. (1982). *The differentiation of society* (S. Holmes & C. Larmore, Trans.). Columbia University Press.
- Luhmann, N. (1990). *Essays on self-reference*. Columbia University Press.
- Luhmann, N. (1994a). *Die Wirtschaft der Gesellschaft*. Suhrkamp Verlag.
- Luhmann, N. (1994b). How can the mind participate in communication? In H. U. Gumbrecht & K. L. Pfeiffer (Eds.), *Materialities of communication* (W. Whobrey, Trans., pp. 371–387). Stanford University Press. (Original work published 1988)
- Luhmann, N. (1995). *Social systems* (J. Bednarz Jr. & D. Baecker, Trans.). Stanford University Press. (Original work published 1984)
- Luhmann, N. (2004). *Law as a social system*. Oxford University Press. <https://doi.org/10.1093/oso/9780198262381.001.0001>
- Lyon, P. (2015). The cognitive cell: Bacterial behavior reconsidered. *Frontiers in Microbiology*, 6, Article 264. <https://doi.org/10.3389/fmicb.2015.00264>

- Lytle, S. R., & Kuhl, P. K. (2017). Social interaction and language acquisition: Toward a neurobiological view. In E. M. Fernández & H. S. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 616–633). Wiley. <https://doi.org/10.1002/9781118829516.ch27>
- Mahner, M., & Bunge, M. (2001). Function and functionalism: A synthetic perspective. *Philosophy of Science*, 68(1), 75–94. <https://doi.org/10.1086/392866>
- Manan, A., Salasiyah, C. I., Rizki, S., & Chairunnisak, C. (2022). Paddy cultivation rituals in South Acèh, Indonesia: An ethnographic study in West Labuhan Haji. *Cogent Social Sciences*, 8(1), Article 2094075. <https://doi.org/10.1080/23311886.2022.2094075>
- Mathews, J., Chang, A. (J.), Devlin, L., & Levin, M. (2023). Cellular signaling pathways as plastic, proto-cognitive systems: Implications for biomedicine. *Patterns*, 4(5), 100737. <https://doi.org/10.1016/j.patter.2023.100737>
- Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. D. Reidel. (Original work published 1972)
- Medzhitov, R., & Janeway, C. A. (2002). Decoding the patterns of self and nonself by the innate immune system. *Science*, 296(5566), 298–300. <https://doi.org/10.1126/science.1068883>
- Miller, A. J. (1971). Consensus and conflict in functionalism: Implications for the study of international integration. *Canadian Journal of Political Science*, 4(2), 178–190. <https://doi.org/10.1017/S0008423900026627>
- Miller, S. L. (1953). A production of amino acids under possible primitive earth conditions. *Science*, 117(3046), 528–529. <https://doi.org/10.1126/science.117.3046.528>
- Minsky, M. (1986). *The society of mind*. Simon & Schuster.
- Minsky, M., & Papert, S. A. (2017). *Perceptrons: An introduction to computational geometry* (Expanded ed.). MIT Press. <https://doi.org/10.7551/mitpress/11301.001.0001>
- Mitchell, A., Romano, G. H., Groisman, B., Yona, A., Dekel, E., Kupiec, M., Dahan, O., & Pilpel, Y. (2009). Adaptive prediction of environmental changes by microorganisms. *Nature*, 460(7252), 220–224. <https://doi.org/10.1038/nature08112>
- Moore, J. L., & Rocklin, T. R. (1998). The distribution of distributed cognition: Multiple interpretations and uses. *Educational Psychology Review*, 10(1), 97–113. <https://doi.org/10.1023/A:1022862215910>
- Moraschi, M., Mascali, D., Tommasin, S., Gili, T., Hassan, I. E., Fratini, M., DiNuzzo, M., Wise, R. G., Salvati, A., Confusion, A., Guerrisi, M., Garaci, F., & Giove, F. (2020). Brain network modularity during a sustained working-memory task. *Frontiers in Physiology*, 11, 422. <https://doi.org/10.3389/fphys.2020.00422>

Morris, I., Seaford, R., Spence, J. D., Korsgaard, C. M., & Atwood, M. (2015). *Foragers, farmers, and fossil fuels: How human values evolve* (S. Macedo, Ed.). Princeton University Press.

Murthy, V. H. (2023). *Our epidemic of loneliness and isolation: The U.S. Surgeon General's advisory on the healing effects of social connection and community*. U.S. Department of Health and Human Services.

Murugan, N. J., Kaltman, D. H., Jin, P. H., Chien, M., Martinez, R., Nguyen, C. Q., Kane, A., Novak, R., Ingber, D. E., & Levin, M. (2021). Mechanosensation mediates long-range spatial decision-making in an aneural organism. *Advanced Materials*, 33(34), 2008161. <https://doi.org/10.1002/adma.202008161>

Nagel, T. (1998). Reductionism and antireductionism. In *The limits of reductionism in biology* (Novartis Foundation Symposium 213, pp. 3–14). Wiley. <https://doi.org/10.1002/9780470515488.ch2>

Nersessian, N. J. (2019). Interdisciplinarity in action: Cognitive ethnography of bioengineering sciences research laboratories. *Perspectives on Science*, 27(4), 553–581. https://doi.org/10.1162/posc_a_00316

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. Wiley.

Nisbett, R. E., & Cohen, D. (1996). *Culture of honour: The psychology of violence in the South*. Westview Press.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

Ormerod, R. (2019). The history and ideas of sociological functionalism: Talcott Parsons, modern sociological theory, and the relevance for OR. *Journal of the Operational Research Society*. <https://doi.org/10.1080/01605682.2019.1640590>

Ostovar, G., & Boedicker, J. Q. (2024). Phenotypic memory in quorum sensing. *PLoS Computational Biology*, 20(7), e1011696. <https://doi.org/10.1371/journal.pcbi.1011696>

Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., & van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, 427(6976), 737–740. <https://doi.org/10.1038/nature02298>

Palikaras, K., Lionaki, E., & Tavernarakis, N. (2015). Balancing mitochondrial biogenesis and mitophagy to maintain energy metabolism homeostasis. *Cell Death & Differentiation*, 22(9), 1399–1401. <https://doi.org/10.1038/cdd.2015.86>

Paniagua, P. (2023, November 6). *Anatomy of externalities*. Centre for the Study of Governance & Society, King's College London. <https://csqs.kcl.ac.uk/anatomy-of-externalities/>

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

Pedersen, T. (2021, December 13). *All about functional magnetic resonance imaging (fMRI)*. Psych Central. <https://psychcentral.com/lib/what-is-functional-magnetic-resonance-imaging-fmri>

Pega, F., Náfrádi, B., Momen, N. C., Ujita, Y., Streicher, K. N., Prüss-Üstün, A. M., et al. (2021). Global, regional, and national burdens of ischemic heart disease and stroke attributable to exposure to long working hours for 194 countries, 2000–2016: A systematic analysis from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury. *Environment International*, 154, 106595.

Petracca, E., & Gallagher, S. (2020). Economic cognitive institutions. *Journal of Institutional Economics*, 16(5), 747–765. <https://doi.org/10.1017/S1744137420000144>

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306(5695), 499–503. <https://doi.org/10.1126/science.1102085>

Pittaluga, G. B., Seghezza, E., & Morelli, P. (2021). The political economy of hyperinflation in Venezuela. *Public Choice*, 186, 337–350. <https://doi.org/10.1007/s11127-019-00766-5>

Polzin, M. (2016). Constitutional identity, unconstitutional amendments and the idea of constituent power: The development of the doctrine of constitutional identity in German constitutional law. *International Journal of Constitutional Law*, 14(2), 411–438. <https://doi.org/10.1093/icon/mow035>

Putnam, H. (1967). The nature of mental states. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). University of Pittsburgh Press.

Rahman, S., & Steeb, D. R. (2024). Unlocking the door to mental wellness: Exploring the impact of homeownership on mental health issues. *BMC Public Health*, 24(1), 3479. <https://doi.org/10.1186/s12889-024-20842-w>

Ranganathan, R., Lee, M.-H., & Newell, K. M. (2020). Repetition without repetition: Challenges in understanding behavioral flexibility in motor skill. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.02018>

Rangaraju, V., Lauterbach, M., & Schuman, E. M. (2019). Spatially stable mitochondrial compartments fuel local translation during plasticity. *Cell*, 176(1–2), 73–84.e15. <https://doi.org/10.1016/j.cell.2018.12.013>

Rawls, J. (1971). *A theory of justice*. Harvard University Press.
<https://doi.org/10.2307/j.ctvjf9z6v>

Rees, T., Bosch, T., & Douglas, A. E. (2018). How the microbiome challenges our concept of self. *PLOS Biology*, 16(2), Article e2005358. <https://doi.org/10.1371/journal.pbio.2005358>

Reid, C. R. (2023). Thoughts from the forest floor: A review of cognition in the slime mould *Physarum polycephalum*. *Animal Cognition*, 26(6), 1783–1797.
<https://doi.org/10.1007/s10071-023-01782-1>

Reuben, A., Schaefer, J. D., Moffitt, T. E., Broadbent, J., Harrington, H., Houts, R. M., Ramrakha, S., Poulton, R., & Caspi, A. (2019). Association of childhood lead exposure with adult personality traits and lifelong mental health. *JAMA Psychiatry*, 76(4), 418–425.

Richmond Federal Reserve Bank. (2024). The origins of the 2 percent inflation target. *Econ Focus*, Q1/Q2.
https://www.richmondfed.org/publications/research/econ_focus/2024/q1_q2_federal_reserve

Robertson, K. (2022). In search of the holy grail: How to reduce the second law of thermodynamics. *The British Journal for the Philosophy of Science*, 73(4), 987–1020.
<https://doi.org/10.1086/714795>

Rogers, J. R., & Lester, D. (Eds.). (2013). *Suicide: A global issue* (2 vols.). Praeger.
<https://doi.org/10.5040/9798216021292>

Rossoll, W., & Bassell, G. J. (2019). Crosstalk of local translation and mitochondria: Powering plasticity in axons and dendrites. *Neuron*, 101(2), 204–206.
<https://doi.org/10.1016/j.neuron.2018.12.027>

Royal Commission into the Robodebt Scheme. (2023). *Report of the Royal Commission into the Robodebt Scheme*. Commonwealth of Australia.
<https://robodebt.royalcommission.gov.au/publications/report>

Sakaguchi, S., Yamaguchi, T., Nomura, T., & Ono, M. (2008). Regulatory T cells and immune tolerance. *Cell*, 133(5), 775–787. <https://doi.org/10.1016/j.cell.2008.05.009>

Sandved-Smith, L., Hesp, C., Mattout, J., Friston, K., Lutz, A., & Ramstead, M. J. D. (2021). Towards a computational phenomenology of mental action: Modelling meta-awareness and attentional control with deep parametric active inference. *Neuroscience of Consciousness*, 2021(2), niab018. <https://doi.org/10.1093/nc/niab018>

Schrödinger, E. (1944). *What is life? The physical aspect of the living cell*. Cambridge University Press.

Schulman, L. S. (2022). *When things grow many: Complexity, universality and emergence in nature* (1st ed.). Oxford University Press.
<https://doi.org/10.1093/oso/9780198861881.001.0001>

Scribner, S. (1975). Recall of classical syllogisms: A cross-cultural investigation of error on logical problems. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. Lawrence Erlbaum Associates.

Sellers, M., & Kirste, S. (Eds.). (2023). *Encyclopedia of the philosophy of law and social philosophy*. Springer Nature. <https://doi.org/10.1007/978-94-007-6730-0>

Sethna, J. P. (2006). *Statistical mechanics: Entropy, order parameters, and complexity*. Oxford University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

Shapiro, J. A. (2021). All living cells are cognitive. *Biochemical and Biophysical Research Communications*, 564, 134–149. <https://doi.org/10.1016/j.bbrc.2020.08.120>

Shapiro, S. J. (2011). *Legality*. Harvard University Press.

Sharma, A., Czégel, D., Lachmann, M., Kempes, C. P., Walker, S. I., & Cronin, L. (2023). Assembly theory explains and quantifies selection and evolution. *Nature*, 622(7982), 321–328. <https://doi.org/10.1038/s41586-023-06600-9>

Shoib, S., Srikanth, S., Javed, S., Gupta, A. K., Swed, S., & Das, S. (2022). Altruistic suicide among Indians during the COVID-19 pandemic — Correspondence. *International Journal of Surgery*, 106, 106933. <https://doi.org/10.1016/j.ijssu.2022.106933>

Siegfried, T. (2000). *The bit and the pendulum: From quantum computing to M-theory — the new physics of information*. John Wiley & Sons.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9(1), 16–20. <https://doi.org/10.1016/j.tics.2004.11.006>

Smythies, J. R., & Koestler, A. (Eds.). (1969). *Beyond reductionism: New perspectives in the life sciences*. Hutchinson.

Solé, R. (2022). Revisiting Leigh Van Valen's "A new evolutionary law" (1973). *Biological Theory*, 17, 120–125. <https://doi.org/10.1007/s13752-021-00391-w>

Spencer, J. P., Perone, S., & Buss, A. T. (2011). Twenty years and going strong: A dynamic systems revolution in motor and cognitive development. *Child Development Perspectives*, 5(4), 260–266. <https://doi.org/10.1111/j.1750-8606.2011.00194.x>

Stack, S. (2004). Émile Durkheim and altruistic suicide. *Archives of Suicide Research*, 8(1), 9–22. <https://doi.org/10.1080/13811110490243741>

Stanford Encyclopedia of Philosophy. (2020). *Multiple realizability*. Retrieved December 30, 2024, from <https://plato.stanford.edu/entries/multiple-realizability/>

Stanford Encyclopedia of Philosophy. (2023). *Functionalism*. Retrieved December 30, 2024, from <https://plato.stanford.edu/entries/functionalism/>

Steels, L. (2011). Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4), 339–356. <https://doi.org/10.1016/j.plrev.2011.10.014>

Steward, O. (2000). *Functional neuroscience*. Springer-Verlag New York.

Stigler, G. J., & Becker, G. S. (1977). De gustibus non est disputandum. *The American Economic Review*, 67(2), 76–90. <http://www.jstor.org/stable/1807222>

Taylor, C. C. W. (1999). *The atomists, Leucippus and Democritus: Fragments, a text and translation with a commentary*. University of Toronto Press.

Telban, B. (2015). Seeing and holding time: Karawari perceptions of temporalities, calendars and clocks. *Time & Society*, 26(2). <https://doi.org/10.1177/0961463X15577273>

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press.

Thibault, P. J. (2011). First-order languaging dynamics and second-order language: The distributed language view. *Ecological Psychology*, 23(3), 210–245. <https://doi.org/10.1080/10407413.2011.591274>

Tho, T. (2020). Mechanical philosophy: Reductionism and foundationalism. In D. Jalobeanu & C. T. Wolfe (Eds.), *Encyclopedia of early modern philosophy and the sciences*. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-319-20791-9_142-1

Tishby, N., Pereira, F. C., & Bialek, W. (1999). The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing* (pp. 368–377). <https://arxiv.org/abs/physics/0004057>

Townsend, M. (2018, November 12). Windrush: 11 people wrongly deported from UK have died, says Sajid Javid. *The Guardian*. <https://www.theguardian.com/uk-news/2018/nov/12/windrush-11-people-wrongly-deported-from-uk-have-died-sajid-javid>

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9)

Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews Neuroscience*, 10(6), 397–409. <https://doi.org/10.1038/nrn2647>

Virtanen, M., Jokela, M., Madsen, I. E., Magnusson Hanson, L. L., Lallukka, T., Nyberg, S. T., et al. (2018). Long working hours and depressive symptoms: Systematic review and meta-analysis of published studies and unpublished individual participant data. *Scandinavian Journal of Work, Environment & Health*, 44(3), 239–250.

von Bertalanffy, L. (1969). *General system theory: Foundations, development, applications*. Penguin University Books.

Walker, M. (2017). *Why we sleep: Unlocking the power of sleep and dreams*. Scribner.

Watanabe, S. (1960). Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, 4(1), 66–82. <https://doi.org/10.1147/rd.41.0066>

Weber, B. H., & Depew, D. J. (Eds.). (2003). *Evolution and learning: The Baldwin effect reconsidered*. MIT Press.

Williams, W. (2020). *Windrush Lessons Learned Review: Independent review by Wendy Williams*. HM Government.
<https://www.gov.uk/government/publications/windrush-lessons-learned-review>

Wilson, E. O. (1998). *Consilience: The unity of knowledge*. Alfred A. Knopf.

Wittgenstein, L. (2009). *Philosophical investigations* (G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans.; 4th ed.). Wiley-Blackwell.

Worthy, D. A., Otto, A. R., & Maddox, W. T. (2012). Working-memory load and temporal myopia in dynamic decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(6), 1640–1658. <https://doi.org/10.1037/a0028146>

Wouters, A. G. (2013). Biology's functional perspective: Roles, advantages and organization. In K. Kampourakis (Ed.), *The philosophy of biology: A companion for educators* (pp. 455–486). Springer. https://doi.org/10.1007/978-94-007-6537-5_21

Zeki, S., & Bartels, A. (1998). The autonomy of the visual systems and the modularity of conscious vision. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1377), 1911–1914.