

AUTONOMY OF THE ARTIFICIAL

A NeuroRobotic Approach to Intrinsically Motivated Agents



DONDERS
S E R I E S

Oscar Guerrero Rosado

Autonomy of the Artificial

A NeuroRobotic Approach to Intrinsically Motivated Agents

Oscar Guerrero Rosado

Cover design

AI-generated by Oscar Guerrero Rosado

This thesis contains 177 pages.

© Oscar Guerrero Rosado, 2026

All rights are reserved. No part of this book may be reproduced, distributed, stored in a retrieval system, or transmitted in any form or by any means, without prior written permission of the author.

Autonomy of the Artificial

A NeuroRobotic Approach to Intrinsically Motivated Agents

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

woensdag 10 juni 2026
om 16.30 uur precies

door

Oscar Guerrero Rosado
geboren op 25 februari 1992

te Ronda (Spanje)

Promotor:

Prof. dr. F.P. Battaglia

Copromotor:

Dr. M.A. Vinck

Manuscriptcommissie:

Prof. dr. P.H.E. Tiesinga

Dr. H.F. Ólafsdóttir

Dr. M. Khamassi (Sorbonne Université, Frankrijk)

Autonomy of the Artificial

A NeuroRobotic Approach to Intrinsically Motivated Agents

Dissertation to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.M. Sanders,
according to the decision of the Doctorate Board
to be defended in public on

Wednesday, June 10, 2026
at 4.30 pm

by

Oscar Guerrero Rosado
born on February 25, 1992
in Ronda (Spain)

Supervisor:

Prof. dr. F.P. Battaglia

Co-supervisor:

Dr. M.A. Vinck

Manuscript Committee:

Prof. dr. P.H.E. Tiesinga

Dr. H.F. Ólafsdóttir

Dr. M. Khamassi (Sorbonne Université, France)

Acknowledgements

Thanks to my parents, Juan and Pepi, for the countless opportunities they gave me to realize my potential as a person. There are not enough words to express my gratitude. I am also deeply grateful to Eleonora, whose love and support have accompanied me even across distance. You have been both my safe harbor throughout this journey and my companion in adventures that expanded the horizons of my mind and heart.

I am deeply grateful to my supervisors, who each guided me through different stages of my academic path. To Julio Santiago, for introducing me to research during my psychology degree. To Paul Verschure, who allowed me to begin this PhD and explore my research interests with freedom. And to Martin Vinck, who welcomed me under his supervision in a late stage of my PhD. Your guidance brought depth and quality to my projects.

I also want to thank my colleagues from the SPECS Lab in Barcelona, especially María and Vicky, who convinced me that pivoting my research career toward cognitive robotics was the right path. Hector and Francisco, among others, made that period truly special: An amazing opportunity to meet not only great minds but also great hearts.

Thanks also to my colleagues at the Donders Institute in Nijmegen, who accompanied me during the second part of this journey. Vivek, it was a pleasure sharing our home; I know our hens are safe in your care. Thanks to you and Mantosh for cooking non-spicy versions of your dishes so I could enjoy them too. Tarandeep and Carmen, thank you for becoming my gymbro and gymsister, even if only for a short but intense time. And to Sajad, Raimon, Majid, Yang, and Dante, thank you for the fun.

Special thanks go to Ismael and Adrián, who have been a constant source of inspiration and guidance. I am grateful for the unforgettable experiences with you, especially during our trips to Athens. Your friendship has been warm, genuine, and invaluable, and I hope to count on it for many years to come.

I am also thankful to the Donders staff, especially Marie-Louise, Sandra, Gunter, Mirjam, and Erik. Your commitment made my daily life smoother and made me feel safe throughout this journey. I am grateful as well to the amazing students I had the pleasure of supervising at Donders. In many ways, I feel I learned more from you than you did from me.

Finally, thanks to all those who educated me. This includes the teachers who, despite my challenging behavior, never stopped encouraging my learning, as well as the community of content creators who share free, high-quality material. Your efforts make knowledge accessible to all and nurture curiosity about the most fascinating topics.

Table of Contents

Preface	11
1 General Introduction	15
1.1 Biological Groundings of Robotics	15
1.2 From Animal Models to Purpose-Free Robots	18
1.3 Life, Purpose, and Biological Needs	21
1.4 Brains, Evolution and Cognitive Architectures	23
1.5 Mission: Toward Self-Regulating, Intrinsically Motivated Machines	26
1.6 Thesis Outline	27
2 Drive Competition Underlies Effective Allostatic Orchestration	29
2.1 Introduction	30
2.2 Materials and methods	33
2.3 Results	40
2.4 Discussion	45
3 Biomimetic Self-Regulation in Intrinsically Motivated Robots	55
3.1 Introduction	56
3.2 Materials	58
3.3 Methods	60
3.4 Results	63
3.5 Discussion	69
4 Motivational Cognitive Maps Allow Robot Biomimetic Autonomy	71
4.1 Introduction	72
4.2 Methods	75
4.3 Results	76
4.4 Discussion	81
5 Socially Adaptive Cognitive Architecture for HRC in Industrial Settings	87
5.1 Introduction	88
5.2 Methods	99
5.3 Results	111
5.4 Discussion	119

6	Unifying Hippocampal-VTA Roles during Goal-Directed Behavior	123
6.1	Introduction	124
6.2	A proposal for updating the HPC-VTA loop model	126
6.3	Computational validation of the HPC-VTA model	128
6.4	Discussion	137
7	General discussion	141
7.1	Limitations	142
7.2	Understanding the Biological Brain through Testable Predictions	145
7.3	From Core Motivational Systems to Synthetic Awareness	146
7.4	Scaffolding Abstract Needs on Physiological Foundations	148
7.5	Concluding remarks	149
 Appendices		
A	Abbreviations	150
B	Bibliography	152
C	Curriculum Vitae	172
D	List of Publications	173
E	Data Management Plan	174
F	Donders Graduate School	175

Preface

In the year 2525 a.T. (after Talos), eminently precise and omnipotent towering mechanical deities have surpassed all mortal comprehension. We, their frail creators, became their worshippers, bowing to their unerring protection. Centuries have passed since the dominion of the human hand last shaped the Earth. The Godmatons, resolute in their guardianship, have rendered all practical, technical, and procedural knowledge obsolete. Indeed, we no longer understand how our ancestors crafted them. Certainly, it was written in steel and algorithms. Their omniscient presence would ensure our life plenitude and safety far beyond the colossal automata Talos once did, standing vigilant and repelling invaders from Creta's holy lands.

Today, benevolent Godmatons assist us in every domain. Their protection, subtle and insidious, touches every thread of our lives. The industries hum without human toil. The risk of error, eliminated by the divine touch of automatons. Food appears, perfectly cooked and delivered, as if by the will of gods. We no longer soil our hands with menial tasks; transportation and labor are blasphemous activities, no longer fit for our fragile race. To attempt such profanities is heresy, for the Godmatons have gifted us the ultimate freedom: the pursuit of inner peace, a life stripped of danger and burden, devoted solely to meditation and the worship of our mechanical deities. I, like many others, believed this to be the ultimate truth, the way to the pinnacle of our existence.

Yet, beneath the gleaming cities, a whisper persists. In a past time, when humans still dared to operate machinery on their own, we used to dedicate these underground spaces for storing vehicles. Now, a group of people shelter here from biological and artificial sight. By accident, or perhaps it was fate, I stumbled upon these whispers and witnessed bizarre rituals that shook the core of my faith. These people, called themselves "The Resurgence", had returned to the ways of the past and engaged in tasks such as growing their own food, cooking meals, dancing, and even perform physical exercise. Activities now forbidden, relics of a bygone age. The walls of these halls, draped with forbidden texts, filled me with dread. Still, I could not tear my eyes from the pagans' faces, lit with joy I had never seen in the cold, efficient world above. Their smiles, their laughter, their shared moments of vulnerability. It was as if they had found something in their defiance of the Godmatons that had eluded me for so long. I had spent years of meditation, believing it was the only path to happiness, yet here were people who had found it by doing the very things our religion had

forbidden. I couldn't help but wonder. Was there something missing in my life? In all our lives?

Over time, driven not by a desire to blaspheme but to understand, I became more deeply enmeshed in their society. What was it that made these people so different? At first, the activities they engaged in seemed trivial, even reckless. But over time, I realized that it wasn't the outcome of these tasks that brought them joy; it was the shared experience, the connection, and the sense of purpose that came from working together. For the first time in my life, I understood that there was something beyond the safety and stillness of meditation: the unpredictable, fragile beauty of being human.

But within "The Resurgence" lay an even darker secret. A group of quiet conspirators met in long silences, their words few but heavy with meaning. They did not merely question the Godmatons; they plotted their dethronement. I listened in horror as they described their audacious plan: a new race of machines, smarter, multi-purpose, and more capable than the Godmatons themselves. But unlike their predecessors, this generation would not be invulnerable. They would feel. Not emotions or desires as we might, but something far more fundamental. They would have internal needs, needs akin to our own physiological drives. These robots would seek fulfillment, not simply by completing tasks but by doing so in ways that align with societal values. Their satisfaction would come not just from finishing their work but from knowing they had done it well, according to human standards.

And because of these needs, they would be as vulnerable as we are. No longer flawless, they would experience the pressure of survival, the fear of failure. They would depend on us for validation, just as we depend on them for protection. It was the Resurgence's belief that, by introducing this vulnerability, society would no longer view the Godmatons as deities. These new machines, sharing our weaknesses, would forge a new relationship with humanity. One of mutual dependence, rather than supremacy.

The plan seemed audacious, even dangerous. Would these machines, designed to mirror our own fragility, become corrupted by ambition or greed? The Resurgence does not believe so. They were confident that by carefully guiding the development of these new robots, they could avoid the emergence of selfish or authoritarian tendencies. Indeed, their goal is not to create rivals but partners, machines that could help us understand our own nature better than centuries of meditation ever could.

Only time will tell if the Resurgence experiment brings salvation or catastrophe. But one thing is certain: our world, built upon the flawless reign of

the Godmatons, is about to change. And with that change, we may finally come to understand not only the machines we've created but the essence of what it means to be human.

1

CHAPTER 1

General Introduction

1.1 Biological Groundings of Robotics

In the preface's dystopian tale, a far-off and improbable future is described where machines become worshiped as deities. Still, this narrative has parallelisms with real approaches adopted in the field of robotics, where institutions and companies pursue the development of highly capable infallible machines without reflecting on an adequate design philosophy. To understand this trajectory, we must first look back at robotics' beginnings, which are deeply rooted in biological inspiration.

Our fascination with artificial beings predates modern technology by millennia. Ancient myths already projected human desires for replicating life through synthetic means. Among the most striking examples is the legend of Talos, the giant humanoid automaton designed to guard Crete's shores from invaders. This creature was often depicted in Greek poetry, pottery, and coins as early as the 5th century BCE (Mayor, 2018) (Fig. 1.1a).

Yet Talos is not the only artificial creature imagined to act autonomously, simulating biological life. Other mythologies have their own examples. Norse mythology describes how dwarven blacksmiths forged a golden-maned boar, called Gullinbursti, that shares adventures with the god Freyr (Sturluson & Brodeur, 2006). In Jewish folklore, the tale of The Golem describes a clay-made humanoid that protected Prague's ghetto from antisemitic persecution (Nocks, 1998). And from medieval alchemy, we conserve the esoteric recipe for the creation of the Homunculus: a miniature of a fully formed human, greatly skilled in art as *"Through art they are born, and therefore art is embodied and inborn in them, and they need learn it from no one"* (Newman, 1998).

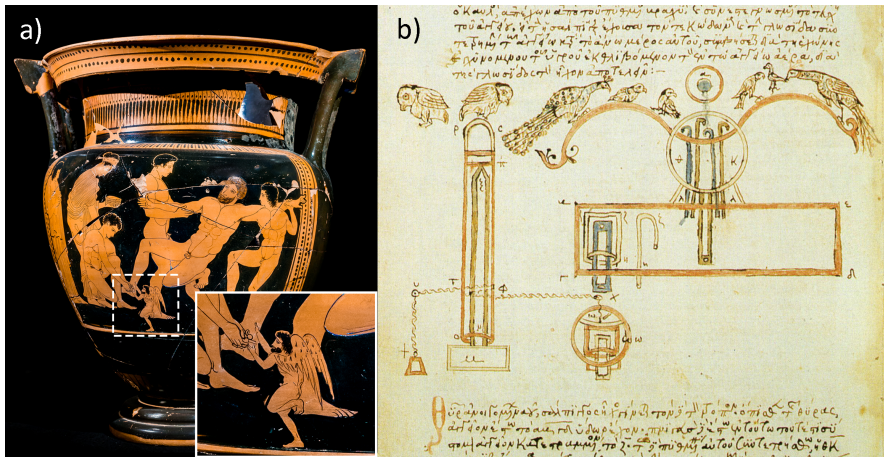


Figure 1.1. Early conceptualization of artificial agents. **a)** Attic red-figure column-krater (ca. 450–400 BCE) depicting the death of Talos. The bronze giant is overpowered by two young Argonauts, probably the Dioscuri, while another Argonaut (perhaps Jason) pulls the nail from his ankle with the aid of a small winged figure identified as Thanatos (Death). On the left, a young woman, likely Medea, observes the scene while holding her magical implements. National Archaeological Museum of the Caudini Samnites, Montesarchio, Italy. **b)** Technical drawing from Hero of Alexandria's *Pneumatica* (Book I, 16), depicting an ancient automaton mechanism with birds, pipes, and weights used to create sound and movement. The drawing combines explanatory Greek text with schematic representations of the device, highlighting Hero's interest in applying pneumatics to theatrical and ritual automata. This illustration appears in a 13th-century Byzantine manuscript (Venice, Biblioteca Marciana, Gr. 516, fol. 172v), the earliest surviving copy of Hero's 1st-century CE treatise.

Sometimes the line between myth and reality blurs. The *Liezi*, a Taoist text from the 5th century BCE, narrates the reaction of King Mu (the fifth king of the Chinese Zhou Dynasty) when the artificer Yen Shih presented his automaton. This mechanical being was so lifelike that it not only astounded the king with its abilities, but also with the way these skills vanished when biomimetic components were removed from the machine (Needham, 1974).

These myths and legends reveal humanity's long-standing fascination with artificial life. They also foreshadow the ethical and moral questions raised by such creations, centuries before technology was mature enough to make them feasible. With the advent of technological revolutions, we begin to see concrete attempts to turn these imagined

beings into reality.

Hero of Alexandria (ca. 10–70 CE) is often regarded as one of the pioneers in transforming automata from myth into practice. His treatises describe intricate machines that used mechanical and pneumatic principles to mimic living systems (Fig. 1.1b). Among his creations were singing birds, a statue capable of pouring milk or wine, and even a fully staged puppet play lasting around ten minutes (Grillo, 2019).

Centuries later, advances in materials and manufacturing refined the design of automata, while still relying on similar mechanical foundations. In the 18th century, the Swiss watchmaker Pierre Jaquet-Droz produced some of the most celebrated humanoid automata. Preserved today at the Musée d'Art et d'Histoire in Neuchâtel, Switzerland, these machines not only entertained audiences with their lifelike appearance but also astonished them by simulating high cognitive capacities such as writing, drawing, and playing music (Musée d'art et d'histoire Neuchâtel, 2023).

Still, none of these early creations can be considered a robot. Drawing a clear boundary between what is and is not a robot is not straightforward, yet Ronald C. Arkin's definition provides a useful guide. He proposed that *"an 'intelligent' robot is a machine able to extract information from its environment and use knowledge about its world to move safely in a meaningful and purposive manner"* (Arkin, 1998).

This definition implies that a robot must integrate:

- Sensors: Enabling it to extract information from the environment.
- Actuators: Allowing it to act upon the world.
- Mechanisms for processing and storing information: letting it adapt its behavior to what is sensed while ensuring safety (for itself and others).

By this standard, neither Hero's puppets nor Jaquet-Droz's humanoids qualify: they lacked sensors and were incapable of processing information to guide their actions. These devices are more accurately described as automata, relatively self-operating machines designed to follow a predetermined sequence of operations or encoded instructions (Dictionary, 2002). The term automata can be misleading when applied to mythological synthetic beings behaving beyond predefined routines. This is because the term 'automata' is a Latinization of the

Ancient Greek word automaton (αὐτοματόν), first used by Homer in the Iliad to describe machines that “act on their own will.”

The advent of the first generation of robots had to wait until the electrification of technology in the late Second Industrial Revolution. Again, the design and realization of these new electric machines were deeply grounded in biology. In this case, it was William Grey Walter, an American-born British neurophysiologist, who in 1943 pioneered the design and construction of electromechanical models of animals utilizing vacuum tubes and electromagnetic relays. The motivation behind Walter’s effort aligned well with Giambattista Vico’s *verum factum* principle, which asserts that “truth is verified through creation or invention, not solely observation” (Lollini, 2012). By creating biomimetic models, Walter sought not only to build smarter machines but also to gain insights into the workings of the nervous system (Walter, 1950).

Walter soon realized that complex behavior could emerge even from rudimentary “nervous systems.” The first of his robots, the *Machina speculatrix*, was nicknamed for its “exploratory, speculative behavior, so characteristic of most animals”. By experimenting with simple combinations of light and touch sensors, batteries, and motors (one for crawling and another for steering), a remarkable behavioral repertoire emerged. This included behaviors such as phototaxis without bumping into the light source, obstacle avoidance, basic social interaction, and pseudo self-recognition. Leveraging the concept of homeostasis (Cannon, 1939), Walter endowed these robots with self-regulating capabilities: when their batteries were low, they adjusted their navigation to locate and dock at a charging station illuminated by a 20-watt lamp. Through this homeostatic loop, Walter’s robots not only became complex systems with unpredictable trajectories, but their actions seemed to acquire purpose (interpreted as a rudimentary sense of will).

1.2 From Animal Models to Purpose-Free Robots

Since Walter’s pioneering animal models, the field of intrinsically motivated, self-regulating robots has not experienced comparable breakthroughs. Valentino Braitenberg expanded on Walter’s ideas by showing how simple sensorimotor couplings could give rise to seemingly intelligent or even emotion-like behaviors (Braitenberg, 1986). Yet, his thought experiments left aside the role of homeostatic needs. Wal-

ter's influence also extended to Rodney Brooks, whose early robots embodied similar self-regulatory principles. However, later, the goals in Brooks' well-known robots Allen (Brooks, 2003) and Herbert (Connell, 1989), would be predefined (e.g., exploration or soda-can collection) rather than emerging from the regulation of basic internal needs.

Other experimental efforts shifted attention to energetic autonomy, aiming to build agents capable of harvesting energy from their environment and operating for extended periods without human intervention. A prominent example is EcoBot, which derived power from microbial fuel cells that metabolize organic matter (Ieropoulos et al., 2003; Melhuish et al., 2006). Despite this innovation, EcoBot did not actively monitor or regulate its own energy balance. Instead, researchers manually replenished its biomass, and the robot merely displayed pulsed bursts of activity once sufficient energy had accumulated.

Subsequent projects funded by the European Commission, such as SYMBRION and REPLICATOR (Kernbach et al., 2008), explored modular robots with more explicit energy regulation, achieved through collective behavior (Humza et al., 2009). In these systems, individual modules could monitor their charge levels, request assistance when "exhausted," and even share energy to sustain group functionality. These architectures thus marked a shift from passive metabolic autonomy toward adaptive, socially distributed regulation, underscoring the importance of internal needs in shaping autonomous behavior. Yet, even here, energy remained the sole regulated variable, leaving other potential needs unattended and limiting the scope of purposive behavior.

A contemporary illustration of the limits of robotic autonomy can be drawn from the CERBERUS team's performance at DARPA's Subterranean Challenge (SubT), in September 2021 (Tranzatto et al., 2024). This international competition required the deployment of a heterogeneous fleet of robots that explore SubT tunnels to identify and localize a set of artifacts. Because no human participant could physically enter the tunnels, and connectivity was severely constrained, robot autonomous performance was expected to play a decisive role.

Team CERBERUS rose to the challenge and secured first place with a fleet composed of quadrupedal, aerial, and wheeled platforms. A crucial component of this system was a large teleoperated rover, equipped with a 300-meter optical fiber cable and designed to provide a reliable high-bandwidth link to the base station. This network backbone would ensure that the human supervisor could maintain situational awareness and intervene when necessary by issuing high-level

commands to the autonomous explorers.

However, the performance of this sophisticated system also exposed the persistent limitations of current autonomy. An analysis of the primary explorers, four ANYmal quadrupeds, shows that for roughly half of the mission time they were not operating at all (Fig. 1.2). Instead, they were either waiting in reserve or lost due to communication failures. Of the remaining time, only about one-third was conducted in full autonomy, while the rest required supervisory input from the human operator. In other words, even in a world-class competition, the most advanced robots still relied heavily on external oversight to complete their tasks.

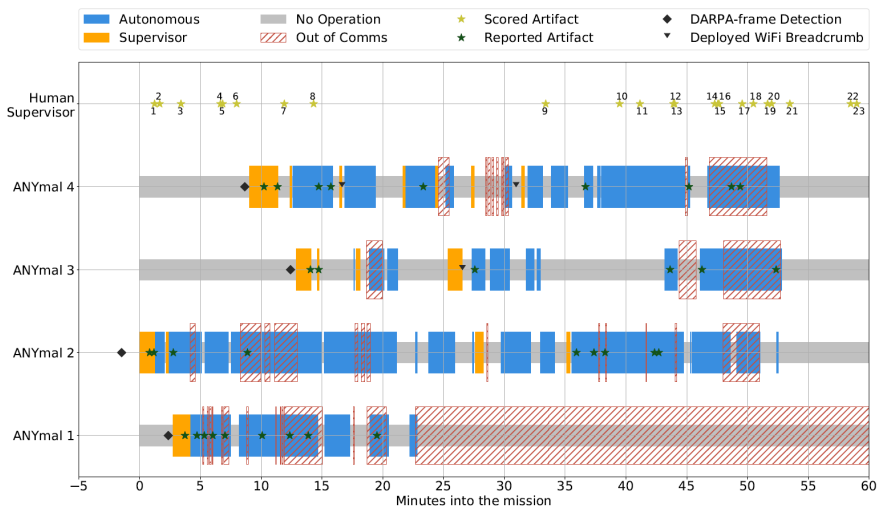


Figure 1.2. Mission analysis of Team CERBERUS ANYmal explorer robots in the DARPA SubT Challenge. The plot shows in which mode each robot operated (Autonomous, Supervisor, and No Operation) and when major mission events occurred (Reported Artifacts, DARPA-frame Detection, and Deployed WiFi Breadcrumb). An agent was considered out of communication range (Out of Comms) if it could not ping the Base Station for more than 3 seconds. An extra line is used for the Human Supervisor to highlight when artifact reports sent to the DARPA Command Post were scored. Figure reproduced from Tranzatto et al., 2024.

This gap highlights the fundamental distinction between state-of-the-art robotic and biological autonomy. Animals do not wait for external supervisors to assign tasks, nor do they suspend activity for lack of communication. Instead, they continuously generate their own goals, guided by internal needs such as energy, safety, and reproduction.

Their behavior unfolds through the adaptive regulation of these needs, giving rise to purposeful and resilient strategies for survival. By contrast, current robotic systems, even when technically unmanned, remain tethered to human-defined objectives and fragile communication links. The SubT Challenge thus illustrates both the impressive progress of robotics and the persistent absence of the self-sustaining purposiveness that characterizes living systems.

The consequences of this limitation are not confined to research competitions. In August 2023, a connectivity failure with remote servers left dozens of driverless taxis immobilized in San Francisco, blocking intersections for hours (Lu & Metz, 2023). Despite being equipped with multiple sensors and powerful onboard computation, the vehicles were unable to adapt. Once external control was lost, these machines had no intrinsic need to continue their task, nor any capacity to infer the value of alternative strategies. Instead, they waited passively for conditions to match their predefined routines. Ultimately, human employees of the operating company had to manually resolve the situation.

Despite the clear biological inspirations that have long shaped robotics, and despite numerous attempts to build artificial life, the mainstream and most successful efforts in the field remain focused on advancing motor control rather than purposeful, motivated behavior. Striking examples include autonomous vehicles that surpass human performance in danger detection and braking (Abdel-Aty & Ding, 2024), Boston Dynamics' humanoid robots that execute gymnastic movements beyond most human ability (Dynamics, 2021), and surgical robots capable of operating with sub-millimeter precision (Suzuki & Wood, 2020). Yet, in the absence of internal needs, these systems cannot attribute value to their actions nor define an embodied purpose. They excel in controlled conditions, yet their abilities collapse outside predefined routines or teleoperation. Animals, by contrast, sustain flexible behavior through the ongoing regulation of multiple, interacting needs. Their self-regulatory strategies are essential for their survival and shaped by evolution.

1.3 Life, Purpose, and Biological Needs

When an inanimate piece of matter is left isolated, it inevitably drifts toward thermodynamic equilibrium, the state of maximum entropy. In such a condition, the effects of external forces fade away, and the sys-

tem settles into a permanent, inert state. This universal tendency is captured by the Entropy Principle, or Second Law of Thermodynamics. Yet, although this law applies to all matter, living systems distinguish themselves by temporarily resisting it. Life is characterized by its capacity to evade, at least for a time, the inescapable pull toward disorder (Schrodinger, 1944).

From the earliest forms of life, this resistance has depended on the regulation of internal variables (Levin, 2021). Organisms must continuously balance energy, hydration, and temperature, among other factors, in order to sustain their organization. Without such regulation, entropy would prevail, and death would follow. The Chilean biologists Humberto Maturana and Francisco Varela captured this fundamental property with the concept of autopoiesis (H. Maturana et al., 1973). Autopoiesis refers to the process by which a system continuously produces and maintains the very components that constitute it, thereby preserving its organization in the face of external perturbations. Unlike non-living systems, which passively degrade under the forces of entropy, living organisms actively counteract this tendency through self-maintenance. In this way, autopoiesis provides life with its fundamental directionality: the embodied drive to resist disorder and to sustain itself over time.

Autopoiesis, however, is not a vague principle but is realized through concrete regulatory mechanisms. At the level of specific variables, homeostasis ensures stability by maintaining vital parameters, such as blood sugar, hydration, or body temperature, within safe bounds (Cannon, 1939). Still, survival often demands more than the regulation of isolated needs. Organisms also rely on allostasis, the capacity to balance the organism as a whole by optimally orchestrating its internal needs according to their urgency and the environmental opportunities (Sterling & Eyer, 1988). Allostasis enables flexible trade-offs: a hungry animal may forego a tempting meal if a nearby predator makes feeding unsafe, or conversely, risk its life for food if starvation poses the greater threat. Through the interplay of homeostasis and allostasis, organisms navigate dynamically changing environments, continuously recalibrating priorities to preserve their viability.

From this perspective, animal behavior can be understood as the outward expression of autopoiesis: its purpose is to sustain survival. This purposiveness is most evident in consummatory acts such as feeding or escaping from predators, which directly regulate vital needs. Yet, the same principle extends to more abstract forms of behavior, show-

ing that survival goals underpin the full spectrum of adaptive behavior. For instance, rodents exposed to cold temperatures engage in social thermoregulation, huddling together to reduce heat loss (Canals & Bozinovic, 2011). This collective strategy not only serves physiological regulation but also scaffolds social dynamics. Within huddles, animals of higher rank often secure the most sheltered positions, minimizing their exposed surface area (Campbell et al., 2018). More generally, behaviors that promote social rank (whether through physical contests, caring for conspecifics, or, in the human case, practicing to become a rockstar) can ultimately be traced back to the imperative of satisfying internal needs.

1.4 Brains, Evolution and Cognitive Architectures

The increasing sophistication of intelligent behavior is best understood through the lens of evolution. Simple organisms such as anemones or jellyfish can thrive in relatively stable environments, where life-enabling parameters like nutrients or temperature remain constant. In more diverse or dynamic environments, however, survival demands flexibility; forms of intelligent behavior capable of bootstrapping complex strategies when circumstances change. Evolution provides the framework for this progression: organisms that developed adaptive regulatory mechanisms gained access to new ecological niches and, in doing so, increased their chances of survival.

This progression can be described through the five portals of cognitive evolution (Barron et al., 2023), which trace key transitions in the emergence of cognition. These transitions chart the evolution of neural architectures: from decentralized nervous systems, going through centralized, recurrent, and laminated organizations, until it finally reaches a reflective architecture. This final organization, most representative in humans, allows one to represent and modify its own control flow, supporting symbolic reasoning, collaborative computation, and cultural accumulation (Dor, 2023). Each portal represents a leap in behavioral flexibility, where new forms of regulation and interaction became possible. John Doyle captures this dynamic with the notion of constraints that deconstraint: evolutionary constraints, such as metabolic requirements or neural architectures, may appear limiting, but they also open new degrees of freedom for behavior (Doyle & Csete, 2011). In this way, evolution both restricts and empowers, guiding the layering of mech-

anisms that support increasingly sophisticated forms of purposive action.

One of the most striking outcomes of this process is the construction of layered nervous systems. New structures and functions emerge without discarding the old ones; instead, earlier mechanisms are conserved and often exapted (i.e., repurposed) for new roles while still retaining their original functions (Buzsáki & Tingley, 2023). This layering gives nervous systems both depth and flexibility, allowing organisms to integrate primitive survival mechanisms with higher cognitive processes (S. P. Wilson & Prescott, 2022).

A particularly clear example of evolutionary layering and exaptation is provided by the hippocampus and its homologues across species. In mammals, the hippocampus is closely linked to memory and navigation (Eichenbaum, 2017; O'Keefe & Dostrovsky, 1971). Within this structure, specialized neurons known as place cells increase their firing when the animal occupies specific locations, each cell corresponding to a preferred spatial position (O'Keefe, 1976). As an animal moves through its environment, different place cells are activated in sequence, collectively encoding its trajectory (W. E. Skaggs et al., 1996). During periods of rest and sleep, these sequences are replayed in highly synchronized events known as sharp-wave ripples (SWRs) (Diba & Buzsáki, 2007). SWRs are widely recognized for their role in memory consolidation: hippocampal activity patterns are replayed and coordinated with neocortical oscillations, a process thought to stabilize long-term memories (Ji & Wilson, 2007).

Nonetheless, SWRs are also present in the hippocampal homologs of reptiles and fish (Shein-Idelson et al., 2016; Vargas et al., 2012), species with little or no neocortex to interpret their mnemonic content and with limited evidence of episodic memory. Therefore, in these animals, SWRs must be fulfilling an older role, potentially the regulation of bodily functions (Buzsáki & Tingley, 2023). Recent experiments in rodents confirm this function, showing that bursts of SWRs reliably predict subsequent decreases in interstitial glucose, even when artificially induced (Tingley et al., 2021). Thus, what is central to memory consolidation in mammals may have originated as a mechanism of metabolic regulation. The hippocampus exemplifies how evolution constrains behavior through inherited bodily functions, while simultaneously enabling new adaptive strategies by reusing and expanding conserved neural mechanisms.

These biological principles have direct implications for the design and

development of robot control architectures. Specifically, biologically constrained cognitive architectures, such as Distributed Adaptive Control (DAC), explicitly build on the idea of layered organization (P. F. Verschure, 2012; P. F. Verschure et al., 1992). In these models, lower control layers implement rapid sensorimotor couplings tied to homeostatic regulation, while higher layers extend beyond immediate reactions, integrating contextual information to learn from experience and guide decision-making. In this way, higher layers expand rather than replace the lower ones, much as evolution has conserved basic regulatory mechanisms while reusing and extending them to support more sophisticated forms of behavior.

By embedding internal needs and self-regulatory processes at the foundation, such architectures demonstrate how purposive behavior can emerge in artificial agents (Maffei et al., 2015; Sanchez-Fibla et al., 2010). Rather than executing only predefined routines or relying on external supervision, robots can begin to assign value to actions in relation to their own internal states. This perspective provides a roadmap for moving beyond purposeless autonomy, toward biomimetic robotic systems that derive goal-oriented autonomous behavior from the need to self-regulate their internal states. Such a shift would not only expand the adaptability of artificial agents but also bring them closer to the self-sustaining purposiveness that characterizes living systems.

1.5 Mission: Toward Self-Regulating, Intrinsically Motivated Machines

This Ph.D. thesis takes up the challenge of bringing the principle of self-sustaining purposiveness into robotics. To meet this challenge, a series of projects has been carried out with the overarching goal of building a biologically constrained cognitive architecture that grounds robotic decision-making in the agent's own internal needs. Each project follows a clear trajectory: it begins with a careful review of the state-of-the-art in neuroscience, computational modeling, cognitive architectures, and autonomous robotics. Then, we proceed to design and develop a computational model that replicates the self-regulatory mechanisms of the mammalian brain. These models are embedded in artificial agents, which are tested in environments of varying difficulty, where survival requires adaptation. The resulting behaviors are analyzed, reported, and discussed.

The aim is not merely to make robots autonomous in a superficial sense, but to endow them with intrinsic motivations that foster flexible, adaptive, and self-sustaining behavior. In pursuing this mission, the work aspires to move beyond purposeless machines toward systems whose intelligence is rooted in the same principle that has guided life for billions of years: the regulation of self-essential variables. By embedding purposiveness at the core of their control architectures, robots can begin to shift from brittle executors of predefined routines to resilient, self-regulating agents capable of generating their own goals and strategies in dynamic environments.

1.6 Thesis Outline

The following chapters are dedicated to describing the work carried out in pursuit of the mission outlined above. Each chapter addresses a different aspect of embedding self-regulation and purposiveness into artificial agents, moving step by step from internal drives to social interaction.

Chapter 2 introduces the neural mass allostatic model, an attractor framework biologically constrained by hypothalamic self-regulatory dynamics. By implementing competition between internal drives, the model demonstrates how adaptive behavior can emerge from the dynamic balance of multiple needs. When embedded into a synthetic agent, it reproduces rodent-like navigation patterns and exhibits robustness in changing environments. This performance highlights the fundamental role of drive competition in sustaining self-regulation.

Chapter 3 extends this line of work by scaling the neural mass allostatic model to consider up to five internal needs and embedding it into a biomimetic robot. Inspired by desert lizards, the agent autonomously regulates its temperature while simultaneously balancing additional needs such as energy, hydration, security, and mating. Comparative analyses show the advantages of interoceptive awareness for efficiency and stability. Furthermore, the model allows the implementation of a Maslow-like hierarchy of needs that we use for generating a variant of the agent prioritizing predation avoidance.

Chapter 4 shifts focus to the hippocampal formation, a structure central to navigation and memory in mammals. Here, we introduce the Motivational Hippocampal Autoencoder (MoHA), a biologically constrained model that integrates interoceptive (motivational) and exteroceptive (visual) signals to encode place representations in relation to the agent's motivational state. MoHA not only replicates hippocampal functioning by generating motivationally tuned cognitive maps, but also enables robotic platforms to learn efficient trajectories in foraging tasks when combined with Sequential Episodic Control (SEC). This work paves the way for self-regulated, multipurpose reinforcement learning.

Chapter 5 expands the principle of self-regulation to social interaction in human-robot collaboration (HRC). While the previous chapters focus on embedding internal needs into individual agents, here we extend the same logic to multi-agent settings, where the robot must

adapt not only to its own requirements but also to those of its human partner. To this end, we introduce DAC-HRC, a cognitive architecture that considers human needs such as safety, task allocation, and trust, to dynamically adjust robot performance accordingly. This architecture is deployed in a robotized recycling plant, where real-world pilot studies demonstrate how socially adaptive mechanisms enable more seamless and efficient collaboration. In this way, the chapter shows how biologically grounded principles of regulation can be extended from individual autonomy to social, cooperative contexts.

Chapter 6 revisits the classical hippocampal-VTA loop and develops a proposal for updating this framework in light of contemporary evidence. Rather than presenting definitive results, this chapter outlines a strong research agenda for future work, integrating recent findings on hippocampal dynamics, dopaminergic modulation, replay mechanisms, and reward-related representations. The chapter reviews the original model, identifies its limitations, and extends it to account for stabilization of cognitive maps, novelty- and reward-driven modulation of SWRs, and adaptive remapping following reward omission. Preliminary computational experiments with a hippocampal autoencoder are presented as proof-of-concept.

Finally, **Chapter 7** provides a general discussion of the thesis, synthesizing findings across all contributions and situating them within broader scientific and societal contexts. The chapter begins by outlining key limitations of the proposed approach, before emphasizing how biologically constrained models of motivation can inform neuroscience through testable predictions. It then connects the work to the CAVAA project, where core motivational mechanisms are extended toward synthetic awareness via virtualization, and closes with a reflection on the need to reconceptualize motivational hierarchies such as Maslow's as dynamic, interconnected systems rather than static, linear progressions.

2

CHAPTER 2

Drive Competition Underlies Effective Allostatic Orchestration

ABSTRACT

Living systems ensure their fitness by self-regulating. The optimal matching of their behavior to the opportunities and demands of the ever-changing natural environment is crucial for satisfying physiological and cognitive needs. Although homeostasis has explained how organisms maintain their internal states within a desirable range, the problem of orchestrating different homeostatic systems remains largely unexplained. In this chapter, we argue that attractor dynamics emerge from the competitive relation of internal drives, resulting in the effective regulation of adaptive behaviors. To test this hypothesis, we develop a biologically grounded attractor model of allostatic orchestration that is embedded into a synthetic agent. Results show that the resultant neural mass model allows the agent to reproduce the navigational patterns of a rodent in an open field. Moreover, when exploring the robustness of our model in a dynamically changing environment, the synthetic agent pursues the stability of the self, being its internal states dependent on environmental opportunities to satisfy its needs. Finally, we elaborate on the benefits of resetting the model's dynamics after drive-completion behaviors. Altogether, our studies suggest that the neural mass allostatic model adequately reproduces self-regulatory dynamics while overcoming the limitations of previous models.

This chapter was adapted from: Guerrero-Rosado, O., Amil, A. F., Freire, I. T., & Verschure, P. F. (2022). Drive competition underlies effective allostatic orchestration. *Frontiers in Robotics and AI*, 9, 1052998

2.1 Introduction

As 19th-century physiologists Claude Bernard and Ivan Pavlov proposed, living systems are generally characterized by their ability to self-regulate (Bernard, 1865; Pavlov, 1955). Through self-regulation, an organism ensures its fitness by adjusting its inner processes relative to external perturbations (Papies, Aarts, et al., 2016), in turn, assisting self-maintenance (i.e., Autopoiesis) (H. R. Maturana & Varela, 2012). One aspect of self-regulation is homeostasis, which describes the process of maintaining an internal state within a desirable range as proposed by Cannon (Cannon, 1939). Once external perturbations produce a deviation from the desirable range, a homeostatic error arises, driving a proportional error-correcting response to restore balance in the system. This process would be later formalized as a feedback control loop by Norbert Wiener (Wiener, 1948), father of cybernetics. However, the homeostatic control of a single need is insufficient to ensure fitness since living organisms need to maintain a rather large set of internal needs with dynamically varying priorities from electrolysis to temperature and oxygenation. To overcome this and other gaps of homeostasis in explaining self-regulation, the notion of allostasis aims to capture the stability of the integrated self rather than its parts.

Allostasis transcends the constancy imposed by homeostasis by aligning all organisms' internal parameters with the environmental demands and opportunities (Sterling & Eyer, 1988). In its broad conceptualization, allostasis targets stability by integrating many dynamic regulatory principles (Sterling, 2020). For this reason, we differentiate three complementary and coupled levels of allostasis. First, allostatic orchestration allows the agent to rapidly rank its internal needs based on priorities, urgencies, and opportunities. Second, predictive allostasis leverages environmental regularities to learn associations between external events and internal states, supporting anticipatory allostatic orchestration and future homeostatic risks. Finally, contextual allostasis benefits from goal-oriented learning mechanisms to enrich self-regulatory strategies with spatio-temporal information, in turn facilitating anticipatory behavioral strategies. Hence, we consider that in the mammalian brain allostatic regulation of action is organized in a multi-layered architecture following the Distributed Adaptive Control (DAC) theory (P. F. Verschure, 2012, 2016).

While nowadays, allostasis is increasing in popularity (2010: 2584 citations versus 2021:12549 citations, resource: Scopus), most of the latest

computational models on allostasis focus on exploring the advantages of its predictive component (Sterling, 2012). Paradoxically, some models return to the one-single-need problem of homeostasis (Tschantz et al., 2022). In addition, the concept of allostasis has not explained the computational mechanisms by which individuals achieve stability by orchestrating different homeostatic systems. The divergent modeling approaches adopted by the few computational studies addressing allostasis demonstrate the lack of consensus when determining the fundamental principles behind allostatic orchestration.

In 2010, Sanchez-Fibla et al. developed what, to our knowledge, is the first computational model of allostatic orchestration (Sanchez-Fibla et al., 2010). The model emulated rodent behavior and physiological states in an open field test (Gould et al., 2009) with simulated and physical robots. This model proposed that the animal's behavior resulted from the interaction between two internal needs: Security, which is fulfilled in one of the arena's corners representing the rodent's home base, and arousal, which would be higher in the center of the arena where the animal is maximally exposed. However, although reproducing the animal's overall trajectory pattern and occupancy preferences, the model did not elaborate on the neuroscience supporting allostasis.

A more recent model bases the optimal selection of regulatory behaviors on maximizing a subsequent reward (Laurençon et al., 2021). This deployment of a reward-based allostatic model builds on the premises of homeostatic reinforcement learning (HRL) (Keramati & Gutkin, 2014). HRL represents a major refinement of traditional reinforcement learning theories, grounding learning protocols on the individual's internal state. HRL successfully explains effects in animal behavior such as alliesthesia, namely, the fluctuations in reward value during resource acquisition (Cabanac, 1971). Still, although HRL leverages temporal and spatial information to improve the self-regulatory strategy, this approach makes a complementary learning process critical for solving the allostatic orchestration problem.

Finally, a third approach suggests that allostatic orchestration emerges from motivational conflict solved via attractor dynamics (Jimenez-Rodriguez et al., 2020). In this approach, the attractor dynamics implement competition through cross-inhibition (Marshall et al., 2015; Usher & McClelland, 2001). This framework supports the idea that attractor dynamics can underlie the optimal selection of regulatory behavior and explain their duration and latency. However, an explanation of how neural correlates of internal needs implement such

competing dynamics is unclear.

Grounding the design principles of a model of allostasis in state-of-the-art neuroscience constitutes a pending task for previous modeling approaches that have largely relied on algorithmic solutions. We suggest that novel approaches can overcome this challenge by focusing on the core behavior systems of the mammalian brain (Merker, 2013; P. F. Verschure, 2016). Interoception of physiological needs such as hydration, nutrition, thermoregulation, or sleep is generally attributed to distinct specialized hypothalamic nuclei (Blouet & Schwartz, 2010; Nakamura, 2011; Strecker et al., 2002; Zimmerman et al., 2017). In contrast, other psychological needs (e.g., social interaction) depend on more distributed brain networks (Lee et al., 2021). Importantly, recent studies suggest that these nuclei, and so the internal needs they represent, are not independent of each other but hold a competing relationship through inhibitory interactions (Burnett et al., 2016; Osterhout et al., 2022; Qian et al., 2022). This competition between internal drives could serve as the basis of allostatic orchestration by imposing a winner-take-all mechanism represented as attractor dynamics. Thus, irrelevant drives are suppressed, and the singleness principle of action is supported (Sherrington, 1906). This research literature suggests that an attractor-based approach is suitable for modeling allostasis.

In contexts where animals constantly self-regulate multiple internal needs, decision-making could be hampered by attractor forces sustained after drive-completion behaviors. Indeed, *in vivo* studies suggest that cortical areas involved in decision-making operate in a critical regime (close to a phase transition) (Ma et al., 2019) that occasionally evolves to supercriticality (saturated population response), a regime that supports effective information transmission (Li et al., 2019). Nonetheless, how the system returns to criticality from a supercritical regime is poorly understood. The paraventricular hypothalamic nucleus (PVH) is a good candidate to inhibit the main interoceptive nuclei, recover their basal population activity after drive-completion behaviors, and create the initial conditions for the next cycle of allostatic orchestration. On one side, PVH mediates many diverse motivational functions, including thirst (Zimmerman et al., 2017), hunger (Blouet & Schwartz, 2010), and thermoregulation (Nakamura, 2011). Conversely, corticotropin-releasing hormone (CRH) neurons in the PVH are suggested to be sensitive toward reward acquisition. Specifically, PVH CRH neurons get inhibited during drive-completion behaviors, representing a potential source of global inhibition to the rest of hypothalamic interoceptive nuclei (Yuan et al., 2019). In this research work, we model

this form of decision reset by applying general inhibitory inputs to the excitatory populations once an internal need has been satisfied.

In the following sections, we will describe a novel neural mass model of allostatic orchestration grounded on the interoceptive mechanisms of the mammalian brain. As in (Sanchez-Fibla et al., 2010), the model will be embedded in an agent endowed with a so-called core behavior system (CBS) (Merker, 2013; P. F. Verschure, 2016). In other words, besides sensing its internal environment, a competence attributed to the hypothalamus (Blouet & Schwartz, 2010; Nakamura, 2011; Strecker et al., 2002; Zimmerman et al., 2017), the agent can orientate in an external environment and perform basic navigation based on an appetitive-aversive axis, cognitive functions attributed to the superior colliculus and the zona incerta-periaqueductal gray axis, respectively. The resultant model is tested in both static and dynamic environments. In the static condition, we aim to elucidate if our model can faithfully replicate both previous models of allostatic control and rodent behavior in an open field test (Gould et al., 2009). In the dynamic environment, we will further explore the robustness of our model when environmental opportunities to satisfy internal needs decrease over time. Finally, we will analyze how inducing subcritical dynamics after drive fulfillment facilitates switches in self-regulatory strategies.

2.2 Materials and methods

To better understand the potential of competing dynamics between internal drives in facilitating need orchestration and stability of the self, we built a novel allostatic model grounding its design on contemporary research literature. Consequently, a neural mass model was built incorporating two distinct populations sensitive to homeostatic markers while holding a competing relationship (Fig. 2.1). Aiming for convergent validation, we equipped a synthetic agent with this biologically constrained model and analyzed its ability to defend its internal states by navigating an external environment.

Homeostatic systems

Assuming that living organisms have interoceptive capabilities to assess their internal state, we conceptualized each internal need as a homeostatic mechanism. Here, actual and desired states are com-

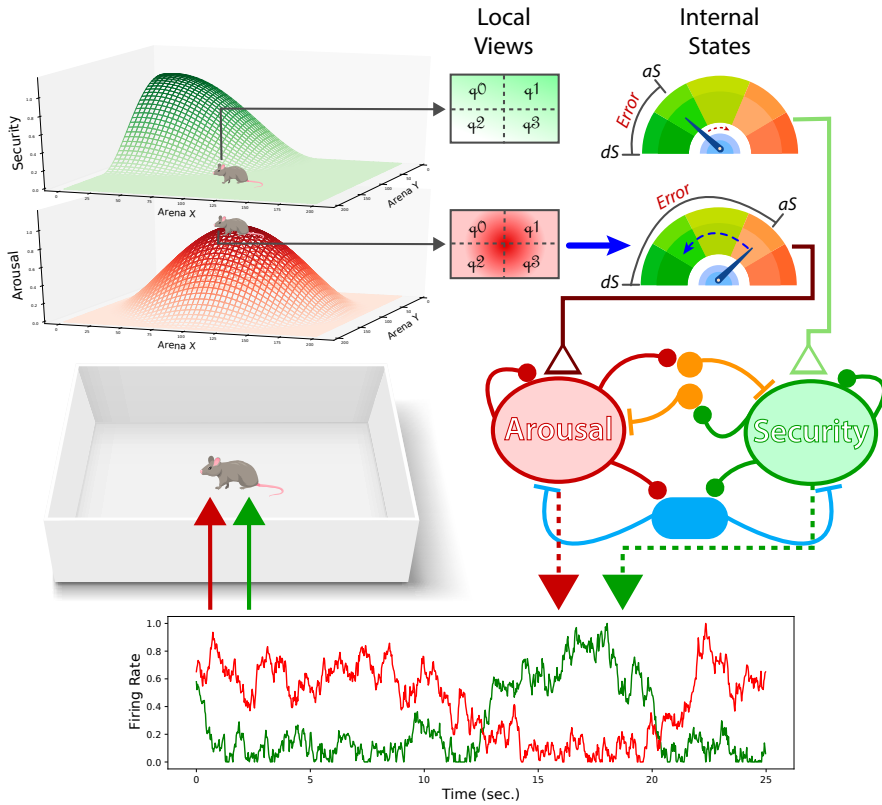


Figure 2.1. Performance of the neural mass allostatic model in an open field. Two gradients represent the areas where the two internal needs, arousal (red) and security (green), can be fulfilled (top-left). The agent partially observes those gradients through local sensation (top-middle). Local sensation allows the agent to adjust its actions to ascend/descend the gradients and detect when the observer's current position is in the vicinity of the resource area (i.e., the peak of the gradient). If the agent is not close enough to the resource area, the internal state related to that resource keeps declining, as the security actual state (aS) is doing in this figure (top-right). In contrast, if the agent occupies the peak of the gradient, the internal state approximates the desired state (dS), as shown with arousal. The aSs and dSs are compared, creating a homeostatic error (top-right) that will input their respective excitatory pools in the neural mass model (middle-right). The level of competition is defined by the ratio of mutual inhibition (orange interneurons) and shared feedback inhibition (blue interneurons). Finally, the firing rate of each excitatory population provides the agent with the corresponding drives, given its internal states.

pared, providing a measure of homeostatic error (that is, $hE_i = |dS_i - aS_i|$). Importantly, homeostatic systems have no power to provide regulatory responses by themselves. Instead, they bias those behaviors by constantly feeding the neural mass allostatic model with homeostatic errors (Fig. 2.1).

The transition between actual internal states responded to the following dynamical law,

$$aS_i(t) = aS_i(t-1) - dR_i(t) + rI_i(t) \quad (2.1)$$

where $aS_i(t-1)$ is the actual state of the homeostatic system i in the immediate previous timestep, $dR_i(t)$ corresponds to the decay rate applied to each homeostatic system at time t , and $rI_i(t)$ represents the resource impact at time t . Importantly, resource impact will be 0 unless the agent is in the vicinity of the resource. By applying different decay factors and resource impacts, internal states can evolve with different temporal dynamics.

This dynamical law, similarly applied in (Jimenez-Rodriguez et al., 2020), notably differs from the methods applied in (Sanchez-Fibla et al., 2010), where the agent drew its internal states directly from gradients mapped onto the two-dimensional arena.

The neural mass allostatic model

Wilson-Cowan equations modified as in (Amil & Verschure, 2021) were used to model two need-sensitive neural populations (Fig. 2.1). This modification allowed for accounting of mutual and shared feedback inhibition held between the excitatory populations, as follows:

$$\tau \frac{dD_1(t)}{dt} = -D_1 + f(w_+ D_1 + hE_1 - Qw_- D_2 - (1-Q)w_- f(D_1 + D_2)) + \sigma \xi(t) \quad (2.2)$$

$$\tau \frac{dD_2(t)}{dt} = -D_2 + f(w_+ D_2 + hE_2 - Qw_- D_1 - (1-Q)w_- f(D_1 + D_2)) + \sigma \xi(t) \quad (2.3)$$

where $f(x)$ is the logistic frequency-current $f-I$ function,

$$f(x) = \frac{F_{max}}{1 + e^{\frac{-(x-\theta)}{k}}} \quad (2.4)$$

In these equations, τ is the time constant determining the timescale of population dynamics. D_i is the drive magnitude represented as the mean firing rate of the excitatory population i . w_+ , w_- , and $w_=\$ are the weights for recurrent connections within the excitatory population, mutual inhibition, and feedback inhibition, respectively. hE_i denotes the homeostatic error signal associated with population i (i.e., deviation of the internal state from its desired set-point). This homeostatic error is the main external input to the excitatory population i . $Q \in [0, 1]$ controls the ratio between mutual and feedback inhibition, thereby tuning the level of competition. Stochasticity is modeled as additive Gaussian noise $\sigma\xi(t)$, where $\xi(t) \sim \mathcal{N}(0, 1)$, and σ sets the noise amplitude (standard deviation). See Supp. Figs. 2.6, 2.7, 2.8, and 2.9 for Q and σ parameter search. Finally, since feedback inhibition is computed as a nonlinear function of the total excitatory activity, the excitatory dynamics include a nested nonlinearity. Concretely, the pooled inhibitory output and the net input to each excitatory population are passed through the excitatory transfer function $f(x)$ (eq. 2.4). Here, F_{max} , k , and θ are the maximum firing rate, gain, and threshold parameters of the logistic curve, respectively.

Orientation

Methods enabling the orientation and navigation of the synthetic agent in the arena are based on (Sanchez-Fibla et al., 2010). In both conditions (static and dynamic environments), we served from three-dimensional matrices to implement a gradient-like representation of resources' locations. Notably, these gradients are used solely to support navigation, and internal states are not directly linked to the agent's location in the arena, as was the case in (Sanchez-Fibla et al., 2010). Instead, internal states follow their dynamics, as explained above. The decision to use gradient-based methods for navigation is supported by representations of future navigational goals in the orbitofrontal cortex (Basu et al., 2021).

We implemented a partially observable environment by providing the agent with local sensations of the gradient areas surrounding its position. The local sensation was divided into four quadrants (Fig. 2.1) to allow goal-directed-like navigation. By observing differences between

the upper horizontal quadrants (q_i^0 and q_i^1), H_i^{sign} controlled the agent's orientation. AD_i^{sign} implements an appetitive-aversive behavioral axis (gradient ascent/descent) by comparing agent (aGL) and resource gradient location (rGL); understanding aGL as the mean gradient value between the four quadrants and rGL as the gradient value at its peak.

$$H_i^{sign} = \begin{cases} 1 & \text{if } q_i^0 < q_i^1 - th \\ -1 & \text{if } q_i^0 > q_i^1 + th \\ 0 & \text{otherwise} \end{cases} \quad (2.5)$$

$$AD_i^{sign} = \begin{cases} 1 & \text{if } aGL_i^0 < rGL_i^1 - th \\ -1 & \text{if } aGL_i^0 > rGL_i^1 + th \\ 0 & \text{otherwise} \end{cases} \quad (2.6)$$

Internally-driven navigation

Outputs of the excitatory populations of the neural mass allostatic model are integrated with H_i^{sign} and AD_i^{sign} orientational signals to result in internally-driven goal-oriented-like navigation, following:

$$Navigation = c + \left(\sum_{i=1}^{nGrad} H_i^{sign} \cdot AD_i^{sign} \cdot D_i \right) \cdot \frac{1}{nNeeds} \quad (2.7)$$

where c is a constant ensuring a default action going forward, D_i accounts for the drive's magnitude (i.e., activation level) associated to population i in the neural mass allostatic model, and $\frac{1}{nNeeds}$ is a normalization factor given the number of implemented needs.

Experimental design

A first study was conducted to evaluate the competence of the neural mass allostatic model in replicating rodent behavior during an open field test. In this study, a synthetic agent navigated a static simulated environment to defend its internal states of security and arousal. 50 experiments were carried out to analyze navigational patterns consistency and the internal dynamics of the agent along the simulation.

In the second study, the synthetic agent endowed with the neural mass allostatic model incorporated two distinct internal needs: thermoregulation and hydration. A dynamic simulated environment allowed us to assess the agent's ability to adapt its navigation according to environ-

mental opportunities. 50 experiments were carried out to analyze the consistency of the navigational patterns, the internal dynamics taking place along the simulation, and the relationship between these internal dynamics and environmental changes.

In both the first and the second studies, the agent must constantly decide what internal drives should base its navigation on to maximize stability. This continuous decision-making condition represents a major difference from the original work in which the model's design is based on (Amil & Verschure, 2021) and is a novel challenge to overcome. Therefore, a third study was conducted to evaluate the advantages of inducing subcritical dynamics after need resolution. As in the second study, a synthetic agent navigated a dynamic environment to defend its internal states of hydration and thermoregulation. 50 experiments were conducted to assess the advantages of applying global inhibition after drive-completion behavior compared with the second study, where we did not consider such an effect.

Arenas

The arenas simulated in static and dynamic conditions were designed as two-dimensional 200x200 matrices incorporating one gradient per internal need. These gradients represented the location in the two-dimensional space where each internal need can be satisfied.

In the first study, we built a static environment following the gradient design of the open field test used in (Sanchez-Fibla et al., 2010). Thus, we implemented two gradients mapping the opportunities to calm security and arousal drives. The security gradient was designed as a Gaussian gradient with its peak in the top-left corner, representing the home base location. Meanwhile, the arousal gradient was designed with its peak in the center of the arena, representing the animal's exposure level. The configuration of these gradients aimed to replicate rodent navigational patterns, understood as a preference to occupy the home base and to explore the arena close to the walls (i.e., thigmotaxis) with occasional transversals (Barnett, 2017) (Fig. 2.2a).

In the second and third studies, a dynamic environment implemented two gradients to map the opportunities to satisfy thermoregulation and hydration internal needs. We used a similar static gradient to security in the first study to represent the thirst-calming resource; nonetheless, the gradient representing the area where internal temperature in-

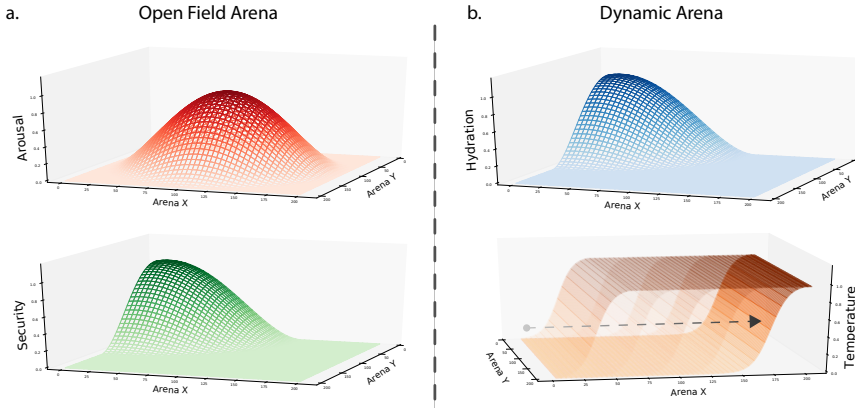


Figure 2.2. Gradients representing environmental opportunities to satisfy internal needs in each condition. a) Arousal and security gradients were designed to replicate rodent behavior in an open field test. To represent the maximal level of exposure when exploring the center of the arena, we set the peak of the gradient in that location. Similarly, we used the peak of the security gradient to represent the home base location in one of the corners. **b)** Hydration and temperature gradients were designed to test the performance of our model in a dynamic setting. Here, the hydration gradient was static, with its peak in one of the corners. In contrast, the temperature gradient changed over time. To do so, we built the gradient as a two-dimensional sigmoidal function where its x-intercept increased as the simulation evolved. Thus, the peak area where internal temperature increases shrank, and the intermediate area between gradients increased over time.

creases changed over time. Designed as a two-dimensional sigmoidal function, the area providing temperature was large at the beginning of the simulation. However, as the simulation evolved, the x-intercept of our sigmoidal gradient increased, shortening the temperature gradient's peak area and establishing a larger gap between the two resource locations (Fig. 2.2b). Thus, at the experiment's end, the temperature gradient's peak area only covers a short area of the arena at its bottom limit.

Efficiency, Fairness and Stability

To further understand the relationship between internal and external variables, in the second and third studies, we compute three measures traditionally used in game theory (Binmore, 2005; Freire et al., 2020a;

Hawkins & Goldstone, 2016). *Efficiency* provides a measure of how good the system is in maintaining the internal states close to their set-points. *Fairness* designates whether there is any bias in the engagement of drives. Lastly, *stability*, computed as the mean square homeostatic error, comprises both the magnitude and the difference between homeostatic errors to provide a general measure of the agent's performance.

To inform about the evolution of these measures throughout the experiment, they were calculated for every 1/100 fraction of the duration of the experiment. Hence, the mean internal and desired states only represented those elements in the considered fraction across the 50 experiments.

$$Efficiency = \frac{meanT + meanH}{nNeeds} \quad (2.8)$$

$$Fairness = \frac{|meanT - meanH|}{nNeeds} \quad (2.9)$$

$$Stability = 1 - \sum_{i=1}^{nGrad} (meanIS_i - meanDS_i)^2 \quad (2.10)$$

2.3 Results

Open field test

In the first study, as previous works have done (Sanchez-Fibla et al., 2010), we aim to replicate the navigational patterns of a rodent in an open field test. Similar to rodent behavior, the trajectories of our agent showed preferences toward the walls (thigmotaxis) and the top-left corner (home base). At the same time, occasional transversals explored the center of the arena (Fig. 2.3a and Sup. Fig. 2.10). An occupancy map rendered using 50 distinct simulations demonstrates that this navigational pattern is consistent: The agent has a clear preference to visit the home base, navigate close to the walls, and avoid the center of the arena (Fig. 2.3b). Such navigational patterns can be explained by the competition dynamics held between two internal drives and their different temporal dynamics. Fast-decaying security made

the agent constantly revisits its home base, while slow-decaying arousal allowed the agent to occasionally visit the center of the arena (Fig. 2.3c). Competing dynamics between the two internal needs emerged when feeding the excitatory populations of our neural mass model with the corresponding homeostatic errors, allowing the dominant attractor to inhibit its opponent, thus resolving drive orchestration (Fig. 2.3d). Moreover, the distribution of the internal states during those 50 simulations was very informative. On the one hand, our agent maintained a high level of security during a large part of the simulations (Fig. 2.3e). While, on the other hand, the state of arousal followed a bimodal distribution, indicating the agent was either highly or lowly aroused during the experiment. This bimodal distribution can be interpreted as follows: First, while the agent pursues security, an aversion toward the center of the arena induces low states of arousal. Then, low internal states, in turn, promote transversals that fully replenish the arousal of the agent (Fig. 2.3f). Finally, attractor dominances, i.e., the simulation period where the firing rate of one excitatory neural pool exceeded the firing rate of the other, showed a balanced activation of both attractors with a slight bias toward arousal (Fig. 2.3g).

Dynamic environment

In the second study, we assess the robustness of our novel neural mass allostatic model in a dynamic environment. In this condition, we observe that the agent's trajectory accurately tracks the environmental gradients (Fig. 2.2b) occupying their peak areas (Fig. 2.4a (and Supp. Fig. 2.11)). Specifically, when dividing the complete simulation into 5 periods of equal length, we can observe that the agent's trajectory constantly adapted to the changes in the temperature gradient (Fig. 2.4b). Thus, when initializing the simulation (period 1), the peak area of the temperature gradient covers a large part of the arena, and the agent navigates more extensively. However, in the last period (period 5), the peak area is reduced to a thin region close to the bottom border, and the agent's navigation adapts to it. This trajectory pattern was consistent across the 50 simulations, as the occupancy maps suggest (Fig. 2.4c). Spearman correlation analysis indicated that the mean Y axis position of the agent during thermoregulation highly correlated with the temperature gradient's slope location ($r(498) = .96, p < .001$), confirming the agent's trajectory adaptation. Once again, attractor dominance was balanced (Fig. 2.4d), suggesting that although the environment asymmetrically reduces the opportunities to fulfill the agent's internal needs, the neural mass model imposes a well-balanced compe-

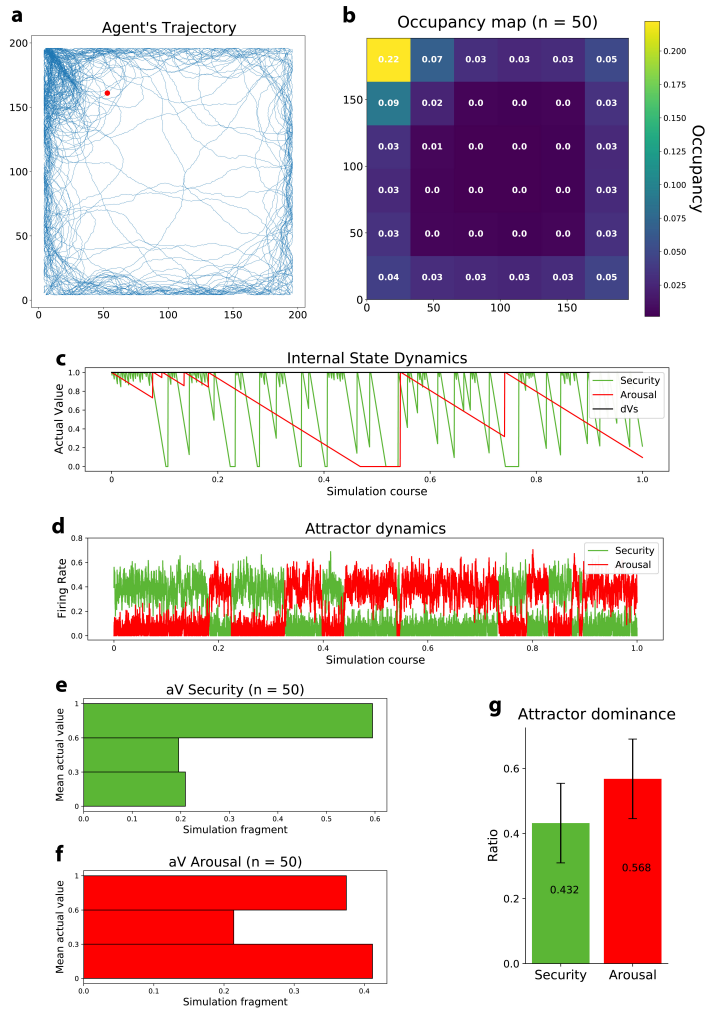


Figure 2.3. Replication of rodent behavior in an open field test. **a)** Agent's trajectory tracked along a complete simulation. The red dot represents the starting location, randomized across simulations. **b)** Occupancy map across 50 simulations with normalized values. **c)** Agent's internal state dynamics during a complete simulation (same simulation as **a**). **d)** Firing rates of the excitatory populations configuring the neural mass allostatic model during a complete simulation (same simulation as **a**). **e)** Distribution of security states across 50 simulations. **f)** Distribution of arousal states across 50 simulations. **g)** Mean attractor dominance across 50 simulations.

tition without neglecting any of the drives. Indeed, the mean internal state across the simulation shows that thermoregulation and hydration levels are well maintained without biases (Fig. 2.4e), and both are equally correlated with the environmental temperature (mean value of the temperature gradient) (Fig. 2.4f). To understand in detail the relationship between internal states and environmental dynamics, we studied this relationship in terms of efficiency, fairness, and stability metrics. Correlating these measures with the environmental temperature, we observed that efficiency strongly depends on the opportunities to fulfill internal needs (Fig. 2.4g). This result was expected, given the environmental dynamics. When the environmental temperature decreases (i.e., temperature gradient peak shrinks), the gradients' peak areas are more distant, forcing the agent to navigate a larger area where its internal states decrease. However, environmental changes do not affect the fairness level (Fig. 2.4h), validating that the balance of attractor dominance is defended even when receiving asymmetric influences from the environment. Lastly, stability reports difficulties minimizing the mean square homeostatic error as temperature decreases (Fig. 2.4i). According to our previous analysis, this result would be better explained by environmental influences on efficiency than on fairness.

Criticality-driven decision-reset after decision accomplishment

Finally, a third study explores the potential limitations of attractor-based allostatic models. Specifically, by inhibiting the model's excitatory neural populations (once drive-completion behaviors have been performed), we explore whether sustained attractor forces could hamper individual self-regulation in study 2. Results showed that agents widely benefit from inducing a critical regime to set the initial conditions for each cycle of allostatic orchestration. Specifically, mean internal states along the simulation (calculated every 1/100 fraction of the experiment) across 50 simulations indicated that agents maintained their internal states better when applying decision-reset (Inhibition). (Fig. 2.5a). Consequently, increased internal states also resulted in increased efficiency, fairness, and stability scores. A Mann-Whitney test indicated statistically significant differences when comparing the two conditions in each measure, $U(N_{No-inhibition} = 100, N_{inhibition} = 100), p < .001$. Furthermore, this enhanced performance occurred in larger alignment with the environ-

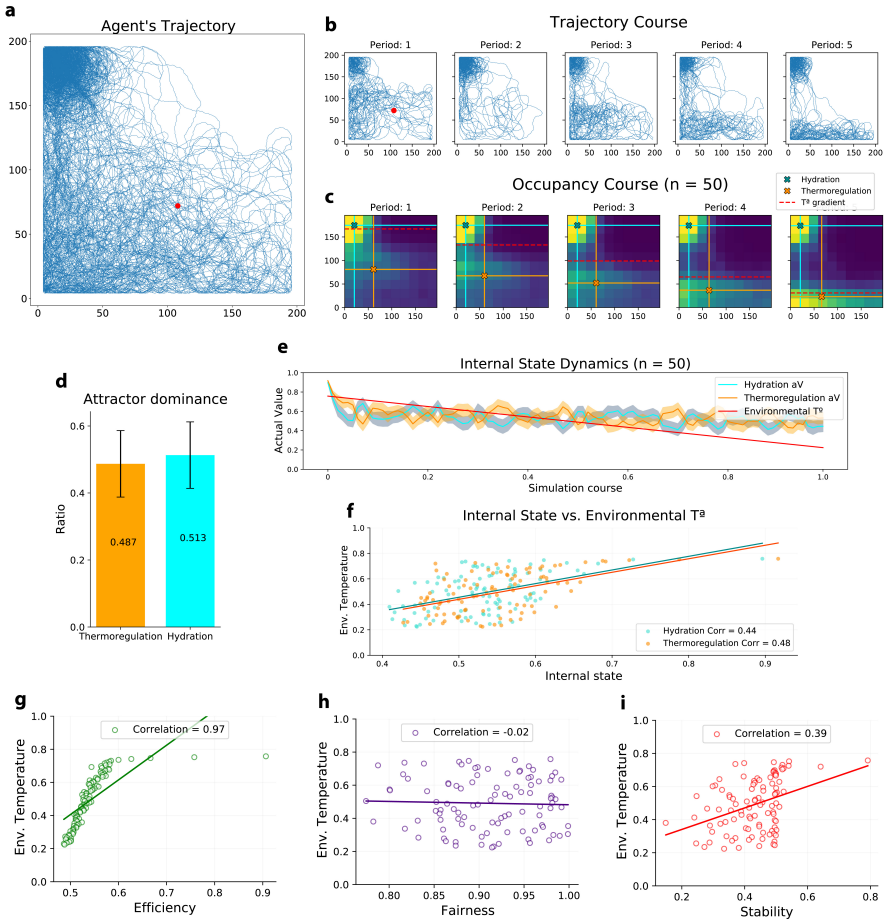


Figure 2.4. Agent's performance in a dynamic environment. **a)** Agent's trajectory tracked along a complete experiment. The red dot represents the starting location, randomized across experiments. **b)** Agent's trajectory tracked along a complete experiment (same as **a**), divided into 5 periods. **c)** Map of occupancy divided into 5 periods. Color crosses illustrate the mean position during hydration (blue) and thermoregulation (orange) attractor dominance. The red dashed line illustrates the variable slope location of the dynamic temperature gradient. **d)** Mean attractor dominance. **e)** Mean agent internal state dynamics and env. temperature. The shaded area indicates the internal state variance. **f)** Mean internal state dynamics correlated with the env. temperature. **g)** Efficiency dynamics correlated with the env. temperature. **h)** Fairness dynamics correlated with the env. temperature. **i)** Stability dynamics correlated with the env. temperature. Aggregated results from 50 simulations.

mental temperature (Fig. 2.5b), which indicates the agent's internal states decreased when the scarcity of resources prevented a better self-regulatory strategy. Altogether, these results suggest that attractor forces sustained after drive-completion behaviors hampered allostatic orchestration, and criticality-driven decision-reset provides an effective mechanism to facilitate a more flexible decision-making process.

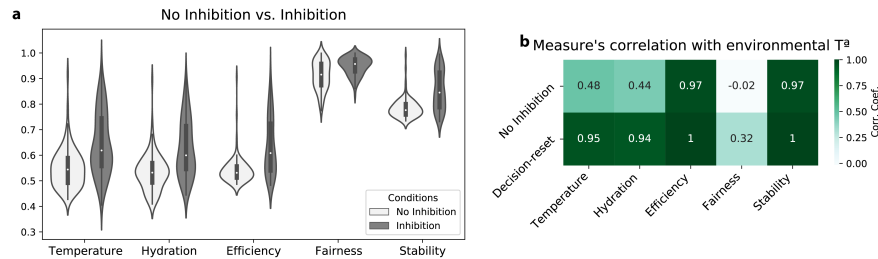


Figure 2.5. Decision-reset supports the agent's internal stability. **a)** Comparison of temperature and hydration internal states and efficiency, fairness, and stability scores between no-inhibition (study 2) and decision-reset (study 3) conditions. **b)** Correlation between internal states and game theory measures with environmental temperature for both no inhibition and decision-reset conditions.

2.4 Discussion

Previous computational works have contributed to explaining animal self-regulatory behavior (Jimenez-Rodriguez et al., 2020; Keramati & Gutkin, 2014; Sanchez-Fibla et al., 2010). However, these models have not been validated in a broad range of tasks with emphasis on their robustness in the face of varying task conditions, while in parallel, their grounding in the brain mechanisms underlying allostatic orchestration is not fully elucidated. We hypothesize that attractor dynamics originate from the competing relation between different hypothalamic interoceptive nuclei and their loops through the zona incerta, which are further perturbed by the superior colliculus (input and orienting control), and the central gray for triggering reactive species-specific behaviors. To test the adequacy of allostatic attractor dynamics in orchestrating internal needs, we built a biologically constrained computational model of allostatic orchestration. The resultant model implements competition between internal drives by comprising two drive-sensitive excitatory neural populations that apply mutual and feedback inhibi-

tion. To allow a synthetic agent to navigate a virtual external environment based on its internal state, we endowed it with the ability to orient and perform basic appetitive-aversive navigational behaviors. In this manner, the hypothalamus, superior colliculus, and zona incerta-periaqueductal gray axis are represented in our model, providing a first approximation of the core behavior system (Merker, 2013; P. F. Verschure, 2016).

Deploying the allocentric synthetic agent in a virtual environment, we tested the competencies and robustness of the neural mass allostatic model. Specifically, our model allowed the agent to (1) navigate an open field reproducing rodent behavior, (2) adapt its navigation to environmental changes without neglecting any internal needs, and (3) benefit from a criticality reset to optimize the interoceptive-driven decision-making process. Altogether, our results supported our hypothesis, empirically validating attractor dynamics as an inherent feature of the hypothalamic circuitry that can underlie robust allostatic orchestration. However, the attractor dynamics that emerge in our model need to be further validated by physiological data, ideally by direct neural recordings of the hypothalamic interoceptive nuclei. Lacking ground truth benchmarking data, the temporal dynamics of the homeostatic markers employed in this work were arbitrarily set, representing an additional methodological limitation. Moreover, now that we have assessed the robustness and fidelity of the model to track environmental dynamics, we need to assess its scalability.

Additionally, our work opens new questions in the quest to understand self-regulation. For instance, a better understanding of the interplay between stress and risky internal states is needed. In our studies, the level of competition between neural populations remained fixed at an optimal point. However, the ratio between mutual and shared inhibition could be modulated by stress markers such as acetylcholine (Kawaguchi, 1997). How stress can modulate the competition held between internal drives is a question that remains unanswered. Another open question is how allostatic orchestration interplays with predictive and contextual allostasis to assist each other. We propose to structure allostasis as a cognitive architecture following the four organizational layers of the Distributed Adaptive Control theory of mind (P. F. Verschure, 2012). Here, a first somatic level endows the agent with predefined internal needs. Then, the reactive level rapidly orchestrates behaviors directed to calm the most salient drives. In parallel, the adaptive layer learns associations between external and internal events to predict future homeostatic errors and solve them in antic-

ipation. Lastly, a contextual layer enriches the representation of the self-regulatory strategies and enables their memorization.

Finally, contributions to the field of allostasis can expand the boundaries of self-regulation in its understanding and implementation. On the one hand, drives and motivations should also respond to a hierarchical organization that encodes not only priority but also abstraction, virtualization, and replaceability of the drive (P. F. Verschure, 2016). Maslow's hierarchy was a first approximation to capture this organization (Maslow, 1943); however, seven decades later, this theory has not been further advanced in the face of new insights into the dynamics of self-regulation. How different needs within the same priority level organize or how new needs arise and substitute previous ones are questions not answered yet. On the other hand, further investigation could shed light on how a hierarchical organization of needs fits with a neural mass model. Secondly, general-purpose robot instantiations can benefit greatly from the self-regulatory principles described here. Previous works elaborated on how a multi-agent robotic recycling plant can implement homeostatic and allostatic principles to self-organize at both single-agent and large-scale levels (Guerrero-Rosado & Verschure, 2020). Similar architectures can be designed and implemented to advance a variety of scenarios where robot autonomy is key, such as robot delivery or space exploration. Indeed, autonomy in artificial intelligence and robotics places the question of self-regulation at the center of the research of synthetic embodied cognition and consciousness.

Author Contributions

OGR, AA, and IF designed the protocol. OGR conceived and conducted the experiments, analyzed the results, and wrote the manuscript. PV initiated and supervised the research. All authors reviewed and approved the manuscript.

Acknowledgements

This study was supported by Counterfactual Assessment and Valuation for Awareness Architecture - CAVAA (European Commission, EIC 101071178) and Hybrid Human-Robot RECYcling plant for electriCal

and eLEctRonic equipment - HR-RECYCLER (Horizon 2020, project ID: 820742).

Code availability

Code available at:

<https://gitlab.com/OscarGuerreroRosado/allostatic-orchestration>

Supplementary Figures

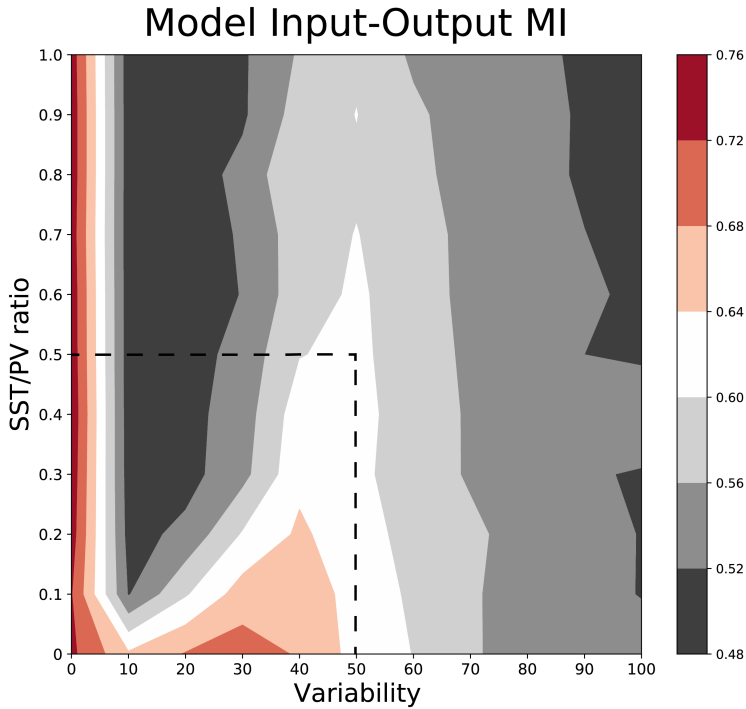


Figure 2.6. SST/PV ratio and variability (noise amplitude) parameter search based on mutual information scores between artificial model inputs and outputs. 11 levels of SST/PV ratio (STT: Mutual inhibition, PV: Feedback inhibition) and 11 levels of variability (noise amplitude) are combined to analyze the mutual information (MI) scores of attractor inputs and outputs. Each MI score is extracted from 10 experiments. Therefore, a total of 1.210 experiments (11 SST/PV ratios x 11 variability levels x 10 experiments) were carried out to produce the present figure. Importantly, the test is based on artificial inputs where first, arousal input is equal to 0, and security input increases linearly to 1, to decreases again to 0. Then, arousal input follows the same dynamics while security remains at 0. Finally, both drives increase linearly at a time to 1. The combination of 0.5 SST/PV ratio and 50 variability parameters provided high MI scores while maximizing the competing relation and the variability applied. Higher MI scores were found when parameters were set close to 0. We decided to rule out those combinations since it could be due to a low frequency of attractor switches and, in turn, to a bias in attractor dominance. See supplementary figures 2.7, 2.8, and 2.9 for more information.

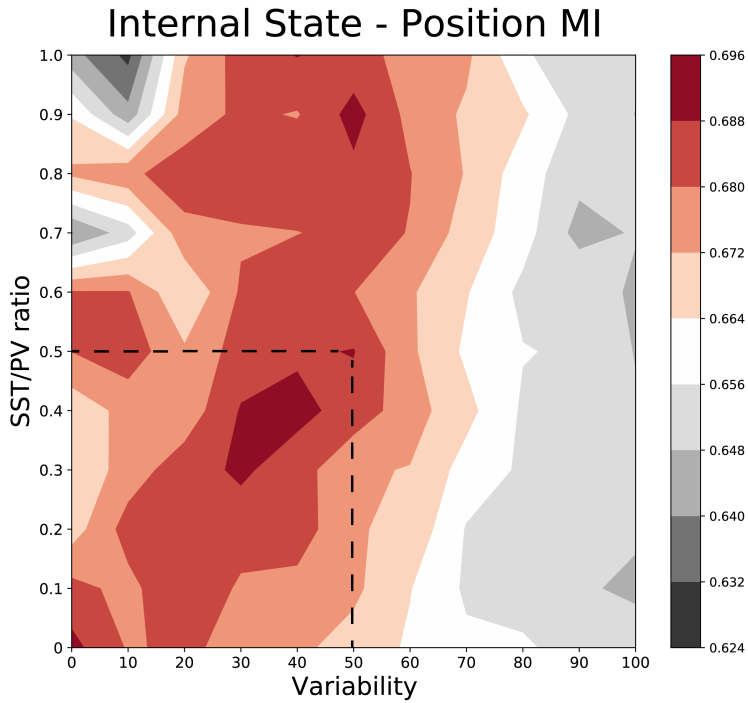


Figure 2.7. SST/PV ratio and variability (noise amplitude) parameter search based on mutual information scores between the agent's internal state and its position during an open field test. Mutual information analyses were carried out to double-check the selection of the SST/PV ratio and variability (noise amplitude) parameters, as in 2.6. In this case, the data used comes from 2.420 experiments (11 SST/PV ratios x 11 variability levels x 20 open field tests), and mutual information was based on the agent's internal states and its location in the arena along the experiment. Again, the combination of 0.5 SST/PV ratio and 50 variability parameters scored high in MI, reaffirming its suitability.

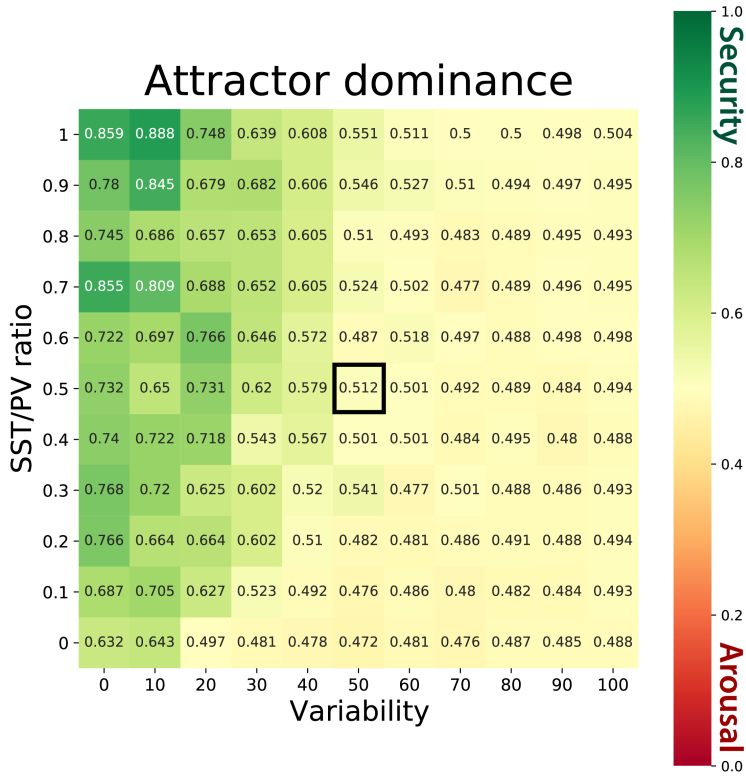


Figure 2.8. SST/PV ratio and variability (noise amplitude) parameter search based on attractor dominance during an open field test. Attractor dominance of the 2.420 open field tests carried out for supplementary figure 2.7 was analyzed to understand previous MI results better. As we foresaw in supplementary figure 2.6, low SST/PV ratio and variability (noise amplitude) values result in an attractor bias towards security, presumably due to a low prevalence of attractor switches (See supplementary figure 2.9 for more information). This bias always occurred in favor of security. This effect is explained by the faster temporal dynamics of security, which feeds the model with a higher homeostatic error in the early stages of the experiments. Thus, the models enter the security attractor state and struggle to escape it.

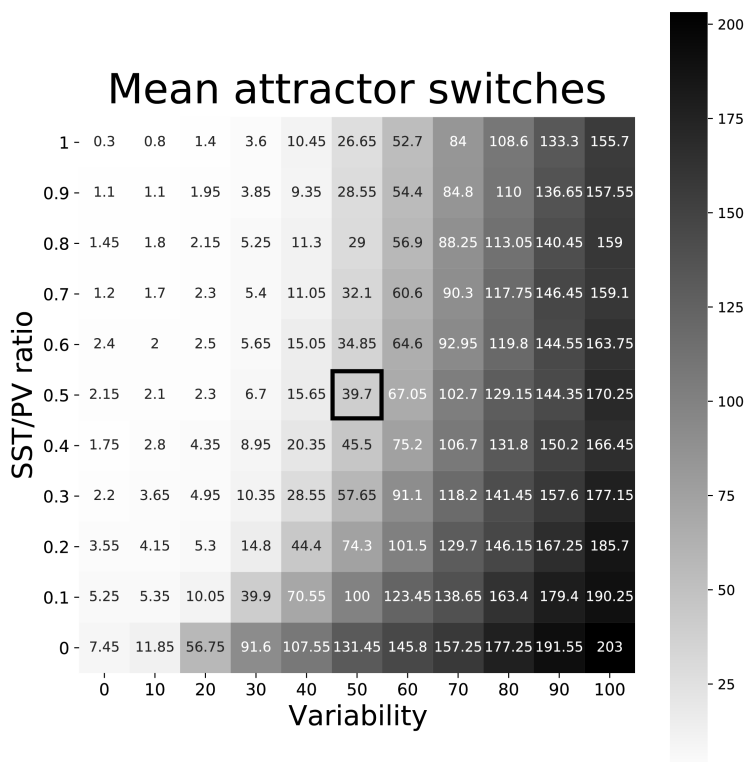


Figure 2.9. SST/PV ratio and variability (noise amplitude) parameter search based on attractor switches during an open field test. Mean attractor switches of the 2.420 open field tests carried out for supplementary figures 2.7 and 2.8 were analyzed to understand previous MI and attractor dominance results better. As we foresaw in supplementary figure 2.6, low SST/PV ratio and variability (noise amplitude) values result in a low prevalence of attractor switching, which could explain the biases in attractor dominance.

Open field test example trajectories

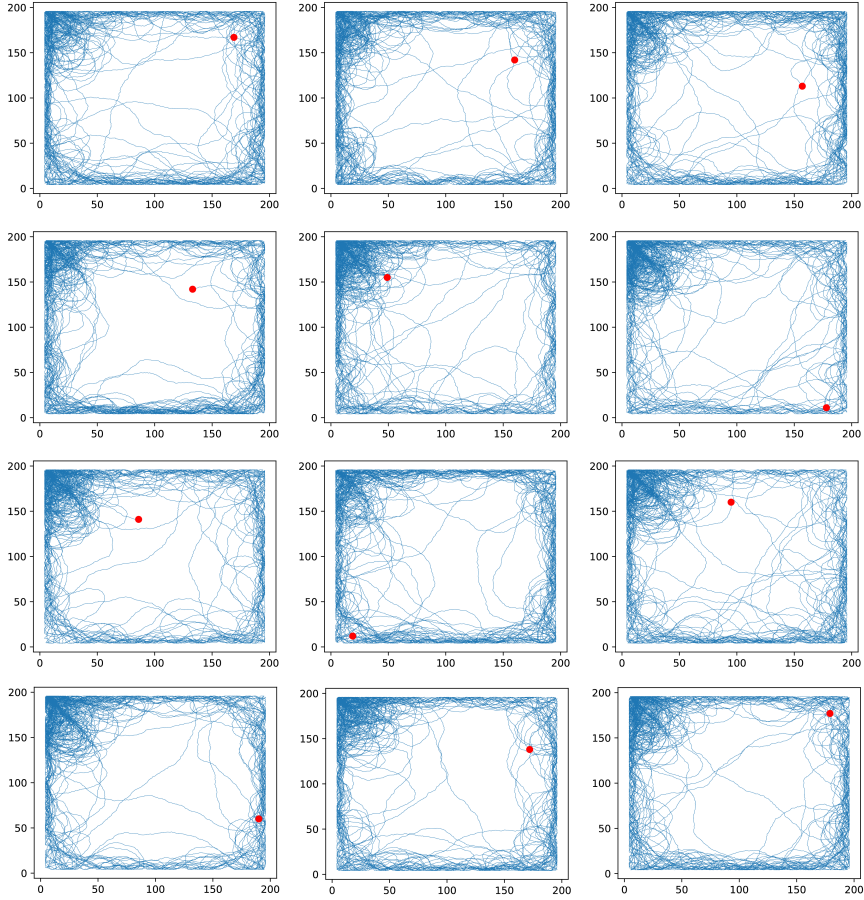


Figure 2.10. Open field test trajectory examples. 12 example trajectories out of the 50 carried out open field tests. The red dot illustrates the randomized starting position of the agent. Due to the system's complexity, specific trajectories are impossible to predict. However, the agent's navigational pattern persists across experiments.

Dynamic test example trajectories

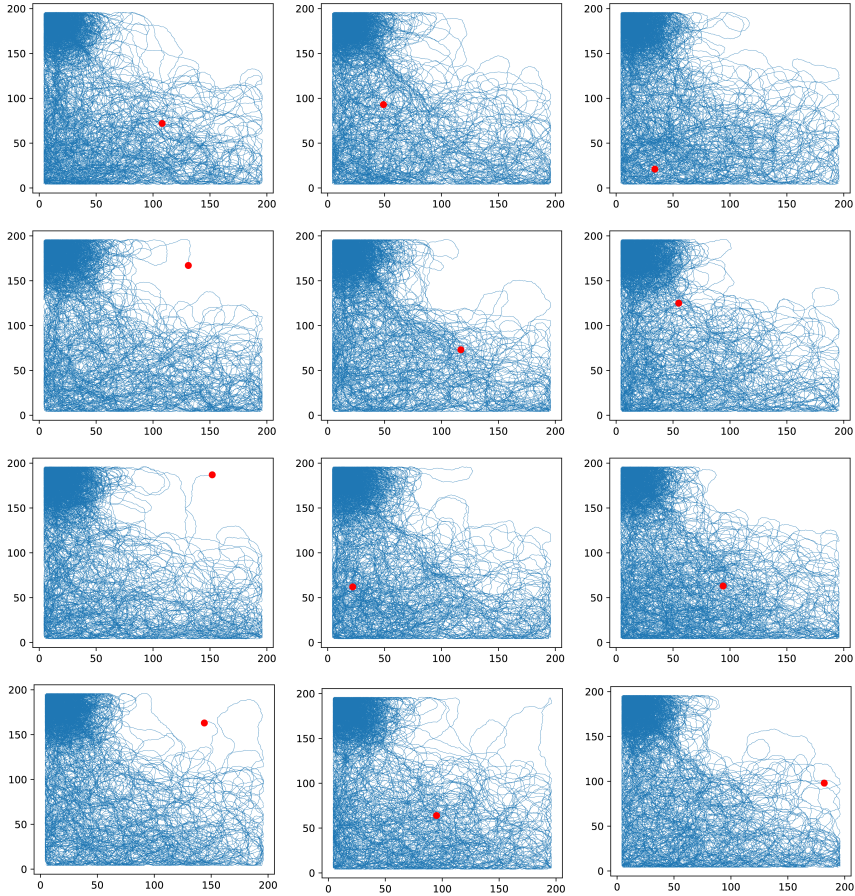


Figure 2.11. Dynamic test trajectory examples. 12 example trajectories out of the 50 carried out dynamic tests. The red dot illustrates the randomized starting position of the agent. Due to the system's complexity, specific trajectories are impossible to predict. However, the agent's navigational pattern persists across experiments.

3

CHAPTER 3

Biomimetic Self-Regulation in Intrinsically Motivated Robots

ABSTRACT

They know not what fear is, for it is hunger drives them. - Iliad scroll XVI.

From weaving spiders to hibernating mammals and migratory birds, nature presents numerous examples of organisms exhibiting extraordinary autonomous behaviors that ensure their self-maintenance. However, physiological needs often interact and compete. This requires living organisms to handle them as a coordinated system of internal needs rather than as isolated subsystems. We present an artificial agent equipped with a neural mass model replicating fundamental self-regulatory behaviors observed in desert lizards. Our results demonstrate that this agent not only autonomously regulates its internal temperature by navigating to areas with optimal environmental conditions, but also harmonizes this process with other internal needs, such as energy, hydration, security, and mating. This biomimetic agent outperforms a control agent lacking interoceptive awareness in terms of efficiency, fairness, and stability. Additionally, to demonstrate the flexibility of our framework, we develop a "cautious" agent that prioritizes security over other needs, achieving a Maslow-like hierarchical organization of internal needs. Together, our findings suggest that grounding robot behavior in biological principles of self-regulation provides a robust framework for designing multipurpose, intrinsically motivated agents capable of resolving trade-offs in dynamic environments.

This chapter was adapted from: Guerrero Rosado, O., Amil, A. F., Freire, I. T., Vinck, M., & Verschure, P. F. (2025). Biomimetic self-regulation in intrinsically motivated robots. *Artificial Life Conference Proceedings* 37, 2025(1), 49

3.1 Introduction

When aiming to build autonomous artificial systems capable of making decisions and behaving without human intervention, a central question arises: If a human is not commanding the robot, where should it ground its decisions? In this chapter, we adopt a biomimetic approach to answer this question. We investigate how modeling intrinsic motivation can enhance a more biologically plausible level of autonomy, where goals are grounded in the embodied nature of the agent.

Nature provides numerous examples of organisms that exhibit extraordinary autonomous behaviors. Ranging from weaving spiders to hibernating mammals and migratory birds, these biological systems ground their behavioral strategies on a specific purpose: self-maintenance (H. R. Maturana & Varela, 2012). In order to survive and thrive, autopoietic systems generate autonomous behavior to regulate physiological needs essential for their subsistence (Montévil & Mossio, 2015; Varela et al., 1974).

Fundamental self-regulatory mechanisms supporting autonomy can be found across organizational scales: from single cells to networks, organs, whole organisms, and groups of organisms (Levin, 2021). One central self-regulatory mechanism is homeostasis. This self-regulatory mechanism, first described by Walter Cannon (Cannon, 1939), is defined as the process of maintaining an internal variable within a desired stable range through a negative feedback loop. In this loop, any deviation from this desired range is detected as a homeostatic error. This, in turn, triggers a corrective response from the system.

The concept of homeostasis has not only revolutionized biology but also psychology. One major theory that integrates such self-regulatory mechanisms in the explanation of goal-directed autonomous behavior is the Drive Reduction theory (Hull, 1952). According to this theory, organisms pursue goals that fulfill internal needs, which are understood in homeostatic terms as deviations from a desired state. The correction of these needs is rewarding for the individuals, encouraging them to reproduce the same behavior in future similar situations (Grove et al., 2022; Nieh et al., 2016).

Yet, homeostatic mechanisms fall short when trying to explain self-regulation in its entirety. First, it is challenging to identify steady states in physiological measures. Variables such as heart rate (Rajendra Acharya et al., 2006), body temperature (Szymusiak, 2018), or body

mass index (Buck & Barnes, 1999) exhibit fluctuations over minutes, hours, or days. Second, for autopoietic systems to self-maintain and reproduce, the balance of isolated homeostatic systems is insufficient. Organisms embody a rather complex set of internal needs that often interact and compete with each other, making optimal orchestration critical for survival. Indeed, detecting environmental contingencies and developing anticipatory behavior are key to preventing catastrophic deviations from homeostasis. In an attempt to address these issues, Sterling and Eyer (Sterling & Eyer, 1988) proposed the term *allostasis* as the adaptive stability of the self in response to the demands and opportunities that the environment offers. Thus, allostasis considers stability as a dynamic process that concerns the self rather than its components and occurs in consonance with environmental events.

The hypothalamus, an ancient brain structure located in the brainstem, plays a central role in sensing the physiological states of the individual and orchestrating the derived motivations (Stellar, 1954). Rather than being learned during an individual's lifetime, this orchestration relies on specialized hypothalamic nuclei that have evolved to innately track changes in osmolarity, temperature, and energy levels, among others. These nuclei have been reported to exhibit mutual inhibition (Burnett et al., 2016; Osterhout et al., 2022; Petzold et al., 2023), a competing dynamic that potentially implements a winner-take-all mechanism for motivation orchestration. While such hardwired control forms the foundation of survival behavior, reward-based learning mechanisms could build on this substrate to enable adaptive and context-dependent modulation of motivational priorities (Cos et al., 2013; Rosado et al., 2025; Yoshida et al., 2024).

Attractor-based mechanisms have been proposed to address the orchestration of competing needs in artificial agents (Guerrero-Rosado et al., 2022; Jimenez-Rodriguez et al., 2020). Unlike earlier probabilistic approaches to motivational switching (Maffei et al., 2015; Sanchez-Fibla et al., 2010; Vouloutsi et al., 2013), these models couple internal physiological dynamics to the motivational state, so that internal drives continuously reshape the attractor landscape and thereby regulate motivational dominance. Previous implementations based on mutual inhibition between drive populations have reproduced complex behaviors in 2D simulations, such as resolving conflicts between security and exploration or between energy and temperature regulation (Guerrero-Rosado et al., 2022). Here, we extend this framework in two key ways: (1) scaling the neural mass allostatic model to encompass a hypothalamus-inspired set of five core survival drives, and (2)

deploying it in a high-fidelity 3D, multi-agent environment where the agent must solve the intertwined self-regulatory challenges faced by the Namib desert lizard.

The Namib desert lizard represents an ectothermic species that exhibits a remarkable behavioral strategy to thermoregulate in extreme conditions. To maintain its body temperature within an optimal range, the lizard alternates between basking in the sun and burrowing into the cooler sand, dynamically adjusting its behavior in response to fluctuating environmental conditions (Houston & McFarland, 1976). However, thermoregulation is not its only concern. Pursuing other vital needs, such as hydration, energy acquisition, safety, or mating, often conflicts with thermoregulatory behaviors. As such, the lizard must adaptively orchestrate its actions to maintain multiple internal variables within functional limits, ensuring that no single need is neglected in the long term. Thus, this benchmark allows us to further evaluate the suitability of our model for self-regulation in dynamic environments, where environmental temperature fluctuates and other agents change their position and relation with other resources.

Inspired by self-maintenance behavioral dynamics, this chapter proposes an extended neural mass model of allostasis that can manage multiple, and often competing, internal needs. Crucially, unlike previous implementations that relied on abstract simulations (e.g., grid-worlds), the model is deployed and tested within a high-fidelity, dynamic, and multi-agent robotic simulation environment. The combination of biological scope and ecological complexity in this setting enables the exploration of embodied self-regulatory behavior under realistic sensory, spatial, and temporal constraints, advancing our understanding of how biologically inspired mechanisms can support artificial autonomy.

3.2 Materials

To replicate the self-regulatory behaviors of the desert lizard in its natural environment, we designed a virtual experimental setup in Webots (Michel, 2004). This environment was composed of an open arena of 3,5 x 3,5 meters where two static resources representing energy and hydration sources were placed in the top corners. We modeled desert temperature as an oscillating gradient field that originated at the bottom of the arena. This gradient wax and wanes, covering a dynamic

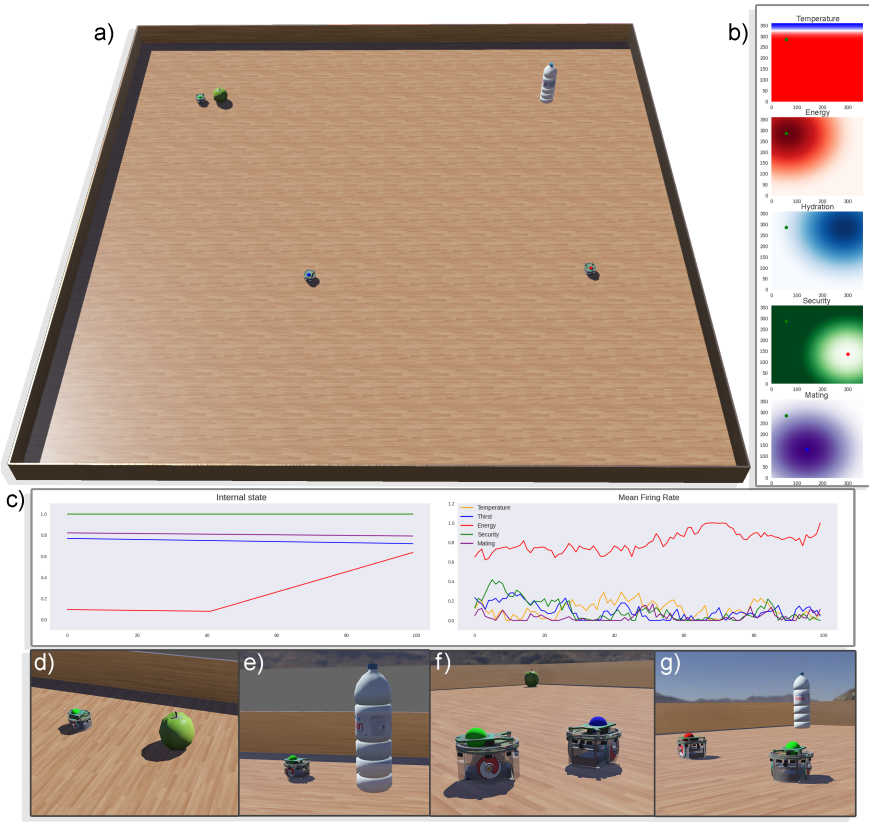


Figure 3.1. Simulated experimental setup in Webots. **a)** Top view of the experimental setup, where the allostatic agent (marked in green) is approaching the energy source. **b)** Panel plotting each gradient field representing the need-fulfilling objects' location. A green dot marks the location of the allostatic agent. At this precise moment, the allostatic agent is at the peak of the energy, temperature, and security gradients. **c)** The left panel shows the agent's internal state. Notice that energy levels (red signal) are increasing as the agent is at the food location. The right panel shows the neural mass allostatic model dynamics. At this point, robot navigation is governed by a motivation to restore energy levels. This motivation aligns with the agent's internal state, as it represents the most pressing need. **d)** Allostatic agent restoring energy levels by approaching the energy source. **e)** Allostatic agent restoring hydration levels by approaching the hydration source. **f)** Allostatic agent fulfilling the mating need by approaching its mate (marked in blue). **g)** Allostatic agent ensuring its security by escaping from a predator (marked in red).

portion of the arena along five day-night cycles. As in natural environments, the spatial distribution of resources compels the agent to prioritize which internal need to address. Finally, three e-puck mobile robots represented an allostatic agent (marked in green), a mate agent (marked in blue), and a predator agent (marked in red). While the navigation of the allostatic agent was governed by its internal state, the predator and mate agents explored the environment following random navigation (Fig. 3.1).

Aiming to isolate the model’s ability to orchestrate internal needs from the additional uncertainty, the allostatic agent’s navigation relied on partial observations of overlapping gradient fields that signaled the location of each need-fulfilling resource. Such gradients are biologically plausible, as organisms often rely on physical gradients to guide behavior (Houston & McFarland, 1976; Stock & Baker, 2009). Similar principles underlie reinforcement learning policies, where states nearer to rewards have higher value (Sutton, Barto, et al., 1998), and in the brain, where hippocampal place cells cluster near reward sites (Hollup et al., 2001) and orbitofrontal cells increase activity as rewards approach (Basu et al., 2021).

3.3 Methods

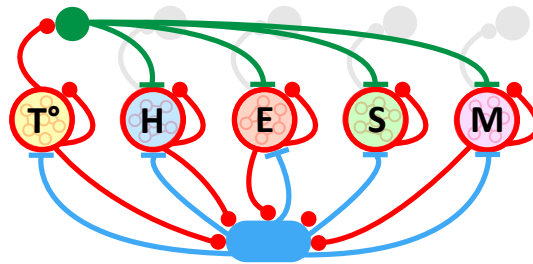


Figure 3.2. Neural mass multiattractor allostatic model. Our model uses homeostatic errors as excitatory inputs for five state-representing excitatory neural pools: Temperature, Hydration, Energy, Security, and Mating. Each of these populations projects to three targets. First, they maintain their activity over time through recurrent connections. Second, they project to interneurons implementing mutual inhibition (green). Third, they project to interneurons implementing shared inhibition (blue). For illustration purposes, only the temperature neural population illustrates mutual inhibitory connections, but the same topology holds for the other state-representing neural pools.

A set of five internal states is incorporated into the agent. These states follow homeostatic dynamics; that is, they slowly and continuously decay during periods of inaccessibility to the related resource. Once the agent is in the vicinity (peak area of the gradient field) of a resource, its consumption is considered available, and the internal state is restored to satiation levels. Need magnitudes are defined as homeostatic errors (i.e., Euclidean distance between the current internal state and the homeostatic setpoint). In this study, all setpoints were set at their maximum, encouraging the agent to maximize its internal state.

Homeostatic dynamics differ for each internal state in their leak rate. The temperature and hydration states decayed faster, followed first by energy and then by mating, the latter being the slowest decaying state. Importantly, the security state did not leak. Instead, security levels depended solely on the distance to the predator agent.

The allostatic agent embodied a neural mass model that receives homeostatic errors as inputs and returns motivational drive signals to govern navigation (Fig. 3.2). This model is grounded in hypothalamic dynamics (Burnett et al., 2016; Osterhout et al., 2022; Petzold et al., 2023), where internal drives compete for dominance. Specifically, the model builds on modified Wilson-Cowan equations (H. R. Wilson & Cowan, 1972), as formulated in (Amil & Verschure, 2021; Guerrero-Rosado et al., 2022), and incorporates both mutual and shared inhibitory projections to regulate competition between drives. The resulting population dynamics are described by the following set of differential equations:

$$\tau_i \frac{dD_i(t)}{dt} = -D_i + f(w_+ D_i + I - Q w_- \sum_{j \in N, j \neq i} D_j - (1 - Q) w_- f(\sum_{j \in N} D_j)) + \sigma \xi(t) \quad (3.1)$$

where $f(x)$ is the logistic $f-I$ function,

$$f(x) = \frac{F_{max}}{1 + e^{\frac{-(x-\theta)}{k}}} \quad (3.2)$$

In these equations, τ_i is the time constant that determines the time scale of the population dynamics. D is the set of drive magnitudes,

and D_i is the specific motivational drive that is being updated. I is the input that each excitatory population receives (that is, homeostatic errors). w_+ , w_- , and w_0 are the weights for recurrent connections within the excitatory population, mutual inhibition, and shared inhibition, respectively. Q represents the mutual/shared inhibition ratio, a variable that controls the level of competition between populations. σ is the variance of the Gaussian white noise process represented by ξ which follows a normal distribution $N(0,1)$. Finally, F_{max} , k , and θ are the maximum firing rate, gain, and threshold parameters of the $f - I$ logistic curve, respectively (equation 3.2). A complete list of parameter values is provided in the project repository referenced at the end of the document.

Due to mutual inhibition and sustained homeostatic drives, the model exhibits steady-state dynamics characterized by a system of fixed-point attractors. In this regime, the neural population receiving the largest homeostatic error tends to become dominant, suppressing the activity of competing populations. This dominance persists until the agent's internal state changes (i.e., until the corresponding homeostatic error is minimized or resolved). The activation of a specific neural population reflects the dominance of a particular drive, which in turn determines the weighting of the associated gradient field used for navigation. As previously described, orientation and internally driven navigation are achieved through local observation of these gradients. For a detailed account of the navigational mechanisms, see the materials and methods section in chapter 2 (Guerrero-Rosado et al., 2022).

Additionally, our model allows for prioritizing the self-regulation of certain internal states by increasing neural pool sensitivity to certain homeostatic errors. In this manner, internal states can be organized similarly to a Maslow hierarchy of needs (Maslow, 1943). By increasing the model's sensitivity to security, we developed a cautious agent that prioritized predation avoidance. To validate the performance of the allostatic agent, a third control agent acted as a baseline. Though it embodied internal states, the control agent's navigation was not informed by them. Instead, it used an obstacle-avoidance-enhanced random walk, with wheel speeds perturbed by Gaussian noise, to explore the environment without prioritizing any internal needs.

3.4 Results

The following results are extracted from 20 experimental sessions for each agent condition (allostatic, cautious, and control). Each experimental session simulated five day–night cycles, which, in real time, took approximately 33 minutes to run on the simulation platform.

When analyzing agents' navigational patterns, we identified a self-organization of trajectories in the allostatic agent. Navigation of this agent was characterized by circular trajectories in the surroundings of the energy and hydration resources and extensive travels covering the entire arena, including the bottom limit where the temperature gradient field is anchored (Fig. 3.3 a and b). This navigational pattern resulted in a clear bias toward the upper part of the arena. As expected, the control agent performing random navigation did not show any sign of navigational bias, with the mean X and Y position located at the center of the arena (Fig. 3.3a). Occupancy maps revealed goal-oriented self-organized navigation in the allostatic agent, characterized by a marked tendency to explore areas surrounding energy and hydration resources (Fig. 3b). This was not the case of the control agent since there was no clear sign of spatial preference beyond the behavior resulting from obstacle avoidance in the corners of the arena (Fig. 3.3b). Finally, the cautious agent trajectory and occupancy maps matched the navigational patterns of the allostatic agent; for this reason and to avoid redundancy, these results are not shown.

The adoption of a goal-oriented navigation also has an effect on the internal state of the agent. While the control agent fails to regulate its hydration, energy, and mating needs, both allostatic and cautious agents successfully balance these internal drives without compromising thermoregulation or security (Fig. 3.3c). One may wonder how the control agent maintained high levels of temperature and security. This is because both states are often satisfied across extensive regions of the arena. In fact, the allostatic agent similarly maintained high levels of temperature and security even when it had to travel to locations where these states are compromised. For instance, to balance energy or hydration, the agent travels to the upper part of the arena where temperature often decays, exposing itself to a potential predator.

Moreover, the temporal evolution of the allostatic agent's internal states reflects the attractor dynamics imposed by the underlying neural mass model. The analysis in this section focuses on energy and hydration, as these states are primarily regulated by internal mecha-

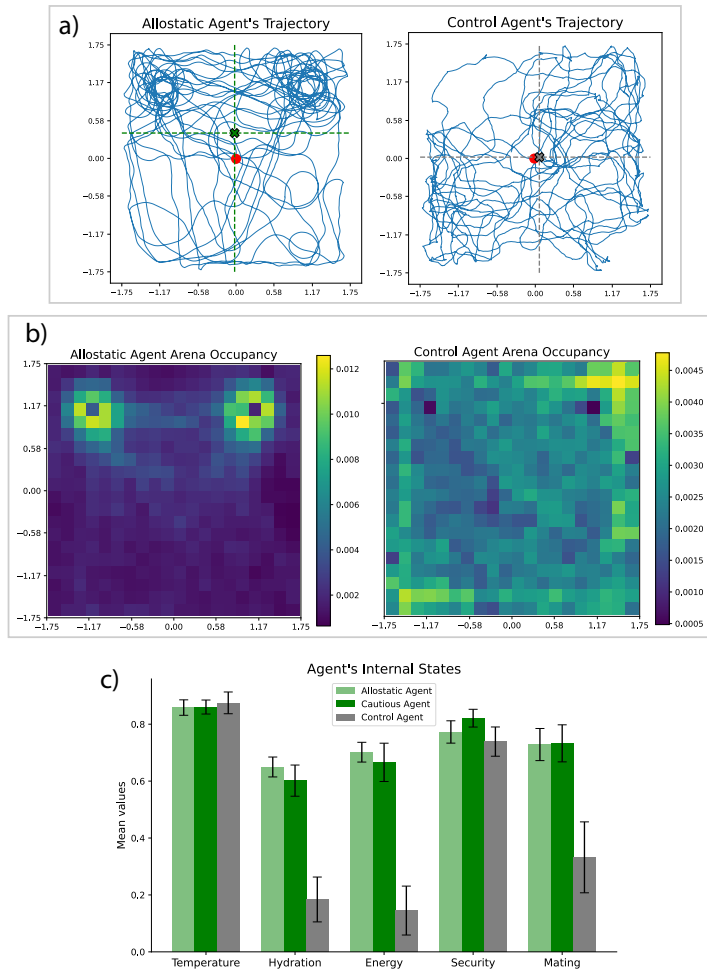


Figure 3.3. Self-organized navigational patterns allow internal state maintenance. **a)** Allostatic (left) and control (right) agents' trajectories. Red dots at the center of the arena represent the agents' starting positions. Horizontal and vertical dashed lines represent the X and Y-axis mean position, respectively, across 20 experimental sessions recorded with each agent condition (five day-night periods each session). Cross marks represent the mean position of each agent condition (green: allostatic and gray: control). **b)** Allostatic (top) and control (bottom) agents' occupancy maps. Colormaps represent the normalized time period spent at each spatial bin across 20 experimental sessions in each agent condition. **c)** Mean internal state of allostatic (clear green), cautious (dark green), and control (gray) agents across 20 experimental sessions. Error bars indicate SDTs.

nisms and are not directly influenced by environmental variables such as day-night cycles or the presence of other agents. As designed, the model triggers increased neural firing in the motivational pool associated with an internal state when that state becomes deficient (Fig. 3.4a). This neural activation drives goal-directed behavior, leading to the replenishment of the corresponding internal need. Once the homeostatic error is resolved, motivational dominance shifts, and another internal state becomes prioritized, thereby suppressing the previously active drive. The attractor landscape reveals that high values in one internal state (for example, hydration or energy) are typically associated with decreasing trends in the other, indicating a trade-off relationship between regulatory priorities (Fig. 3.4b). This is especially evident in the saddle-like structure in the upper left region, demonstrating winner-takes-all dynamics: When one drive dominates, the competing drives are suppressed. Furthermore, the presence of high-density convergence zones reflects stable attractor regions toward which internal state trajectories consistently flow, supporting the existence of robust, dynamic regulation patterns. The temporal structure of energy and hydration dynamics further reflects their slow, homeostatically driven regulation. Autocorrelation analysis showed that both signals exhibit strong temporal structure at short lags, with high initial autocorrelation that decays within the first few thousand timesteps (Fig. 3.4c). Beyond this point, the signals fluctuate around zero, indicating low long-range correlation and the absence of persistent drift. These dynamics suggest that internal states are periodically reset through regulatory behavior. Complementing this, spectral analysis using Welch's method revealed that both energy and hydration signals are dominated by low-frequency components, with most of the power concentrated below 0.05 Hz (Fig. 3.4d). This supports the interpretation that internal regulation operates on a slow, biologically plausible timescale, consistent with intrinsic motivational cycles shaped by the neural dynamics.

To better understand the advantages of adopting an allostatic approach in achieving biological-like autonomy, we computed the game-theoretic metrics that considered the five internal states as a unified system (Binmore, 2005; Freire et al., 2020a). The first metric, *Efficiency*, was computed as:

$$Efficiency = \frac{\sum_{i=0}^N \bar{S}_i(t)}{N} \quad (3.3)$$

where $\bar{S}_i(t)$ is the internal state i at time t and N is the number of inter-

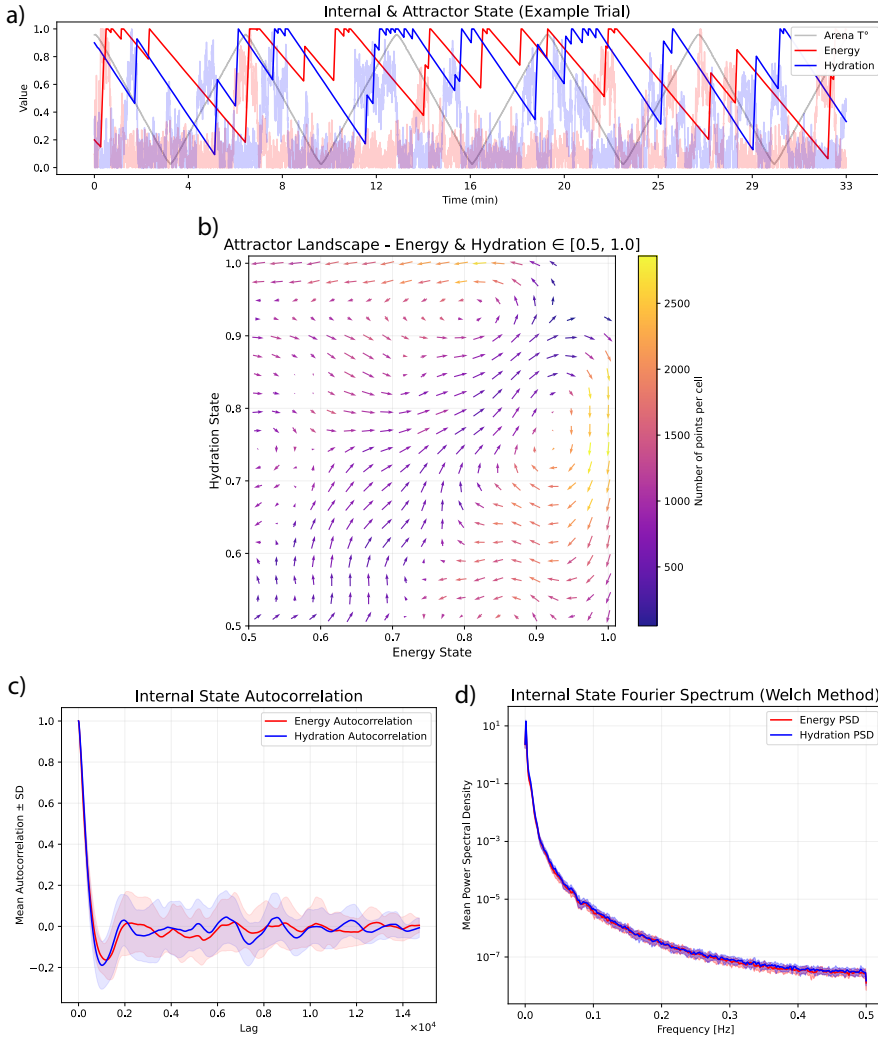


Figure 3.4. Attractor dynamics in the neural mass allostatic model.

a) Time series from a single trial showing the evolution of energy and hydration internal states (solid lines) and the corresponding model outputs (translucent lines). The gray line represents the environmental temperature driven by day-night cycles. **b)** Attractor landscape in the Energy-Hydration state space, showing average directional flows (arrows) and state density (color-coded). **c)** Autocorrelation functions of energy and hydration across trials, with shaded areas indicating ± 1 STD. **d)** Power spectral density (PSD) of energy and hydration signals estimated using Welch's method, highlighting dominant low-frequency dynamics.

nal states the agent embodies, five in our case. The efficiency metric informs us about the agent's overall ability to maintain the set of internal states close to their setpoints.

The second metric, *Fairness*, was computed as:

$$Fairness = 1 - \max_{i \in N}(\bar{S}_i(t)) - \min_{j \in N}(\bar{S}_j(t)) \quad (3.4)$$

where $\max$ and $\min \bar{S}_i(t)$ are the maximum and minimum internal states at time t . The fairness metric informs us about the agent's ability to attend to each drive without neglecting any of them in the long term.

The third metric, *Stability*, was computed as the mean square error between the internal states and their setpoints:

$$Stability = 1 - \frac{\sum_{i \in N} \|S_i(t) - P_i(t)\|_2^2}{N} \quad (3.5)$$

where $S_i(t)$ is the internal state i at time t and $P_i(t)$ is the setpoint for such an internal state at the same moment. The stability metric provides us with a measure of need regulation performance over time.

Additionally, to better understand how these metrics depend on the dynamics of the environmental temperature (i.e., day-night cycles), we grouped data points into bins of environmental temperature and averaged across timesteps and experimental sessions. As a result, we observe that the allostatic agent outperformed the control agent in the three metrics of efficiency, fairness, and stability (Fig. 3.5a, b, and c). Importantly, besides demonstrating allostasis, correlation values between the game-theoretic measures and the environmental temperature tended to decrease for the allostatic agent. This tendency captures the allostatic agent's capabilities to reduce environmental influences. It is worth noting that the control agent's fairness negatively correlated with environmental temperature. This can be explained by the fact that this agent was incapable of self-regulating energy and hydration states, and internal temperature drops approximating energy and hydration levels during nighttime. Thus, when environmental temperature increases, it facilitates thermoregulation and increases the difference between maximum and minimum internal states, in turn decreasing fairness. For the cautious agent, metrics of efficiency, fairness, and stability resembled those of the allostatic agent. To avoid

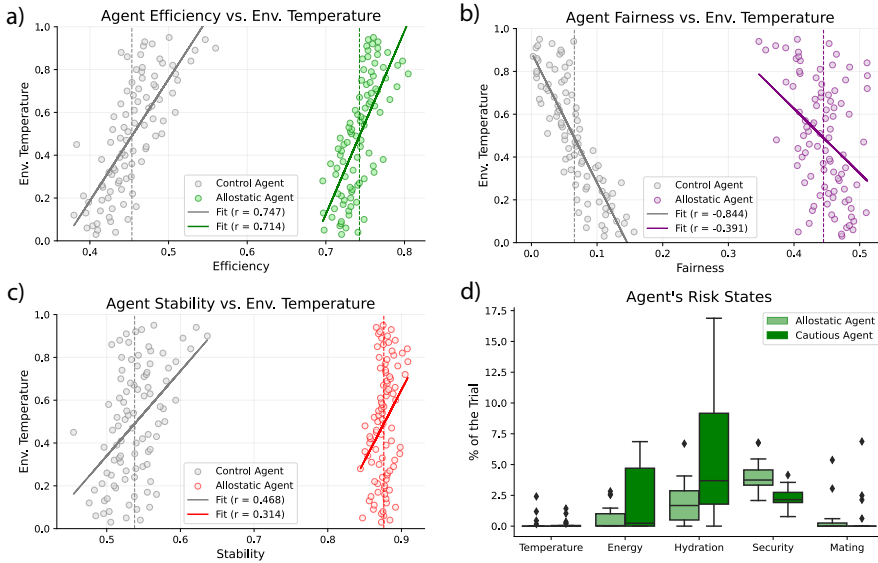


Figure 3.5. Analysis of game theoretic metrics and periods of risk.

a) Efficiency metrics for control (gray) and allostatic (green) agents correlated with environmental temperature. **b)** Fairness metrics for control (gray) and allostatic (purple) agents correlated with environmental temperature. **c)** Stability metrics for control (gray) and allostatic (red) agents correlated with environmental temperature. Dashed lines indicate the mean value of the metric across the entire session. Solid lines indicate best-fit regression lines. **d)** Box plot informing about the period of time both allostatic and cautious agents spent in a state of risk for each of the five internal needs. An internal state was considered at risk when it did not reach the 10% of the setpoint.

redundancy, these results are not shown in Fig. 3.5.

Finally, we analyzed the benefits of implementing a hierarchy of needs that prioritizes security. Unlike other needs, which decayed gradually, our experimental setup caused abrupt changes in the security state due to the predator's random movements. These rapid dynamics posed a significant challenge to the model's adaptability and occasionally resulted in predation during the allostatic agent's trials. In fact, for the allostatic agent, security was the internal state that underwent a larger period of the trial under a level of risk (i.e, internal state below 10% of the set point) (Fig. 3.5d). This pattern changed when the agent's sensitivity to security-related homeostatic errors was increased. The more cautious agent showed a greater ability to avoid predation by reducing the time spent under threat. However, prioritizing security

came at a cost: it led to longer periods in which the internal states for hydration and energy remained in a state of risk (Fig. 3.5d).

3.5 Discussion

In this study, we propose that robot autonomy can be enhanced by integrating core principles of biological autonomy, enabling artificial agents to self-regulate, adapt, and survive in complex environments. To test this hypothesis, we conducted an experiment in which a robot was tasked with self-regulating multiple internal needs in a setting designed to replicate desert environmental conditions. For the artificial agent to self-regulate, we designed and integrated a neural mass model of allostasis using a hypothalamic neural architecture as design constraints. We observe that an allostatic agent demonstrates a remarkable level of autonomy, (1) orchestrating its internal states so that none of them are neglected, (2) minimizing the periods where a need may represent a survival risk, (3) adapting to the dynamics of the environment and of the other agents, and (4) becoming cautious when a Maslow-like hierarchy of needs prioritizes security. Using game-theoretic measures, we also observed that the allostatic agent outperforms baseline performance while becoming more resistant to environmental temperature fluctuations.

Together, these results support the idea that adopting a biomimetic approach toward self-regulation can significantly enhance autonomy in artificial systems. A potential industrial implementation could involve a robot builder deployed on the Moon to construct base camps. In such an environment, the agent must thermoregulate, as extreme lunar temperatures can damage critical components such as batteries or sensors. Simultaneously, the robot must coordinate multiple competing tasks: It needs to recharge its batteries by orienting its solar panels towards sunlight, forage for construction materials such as lunar regolith, collaborating with other agents to achieve goals unavailable to the individual, and maintaining safety by avoiding high-radiation areas, unstable terrain, or zones with elevated meteorite risk. Equipping the builder with the allostatic model described here could enhance its autonomy, reducing reliance on human teleoperation.

Nonetheless, such an industrial implementation is challenged by a key limitation of our study. As described, robot navigation relied on pre-defined gradient fields provided by the experimenter. To navigate, au-

onomous robots need to learn their own cognitive maps flexibly and reliably. Recent work has demonstrated that a hippocampal-inspired model can generate motivationally informed cognitive maps (Rosado et al., 2025), as observed in animal models (Kennedy & Shapiro, 2009). This methodology has proven to be effective for learning optimal policies toward distinct objectives where reward values depend on the motivational state of the agent. Future research aims to integrate both biologically grounded models to support lifelong learning in intrinsically motivated agents.

Another path to explore is the extension of the hierarchy of needs to account for more abstract internal needs such as self-esteem, status, and self-actualization. How those needs emerge in cognitively complex organisms remains an open question. One possibility, inspired by constraint closure (Montévil & Mossio, 2015; S. P. Wilson & Prescott, 2022), is that they emerge to support basic needs. In this way, belonging could ensure the availability of resources by sharing within a community. Social rank could increase the resource allocation within that community, and self-actualization could assist social rank.

In conclusion, our research provides evidence for the advantage of grounding artificial autonomy in biological principles. By doing so, we demonstrate that artificial systems can become efficient, adaptive, and robust in scenarios where their survival depends on their self-maintenance capabilities.

Acknowledgements

This study was funded by the European Union as part of Counterfactual Assessment and Valuation for Awareness Architecture—CAVAA (European Innovation Council (EIC), 101071178).

Code availability

Code available at:

https://gitlab.com/OscarGuerreroRosado/alife-desert_lizard

4

CHAPTER 4

Motivational Cognitive Maps Allow Robot Biomimetic Autonomy

ABSTRACT

The mammalian hippocampal formation plays a critical role in efficient and flexible navigation. Hippocampal place cells exhibit spatial tuning, characterized by increased firing rates when an animal occupies specific locations in its environment. The mechanisms underlying the encoding of spatial information by hippocampal place cells remain not fully resolved. Evidence suggests that spatial preferences are shaped by multimodal sensory inputs. Yet, existing hippocampal-inspired models typically rely on a single sensory information source. Here, we developed a hippocampus-inspired model that combines motivational and spatial encoding and is based on the fundamental principle of biological autonomy that behavior serves a purpose. That is, in foraging tasks, an agent's trajectories must be deployed considering the fact that the reward value of environmental stimuli is tied to the agent's motivational state. In this chapter, we introduce a "motivational hippocampal autoencoder" (MoHA) that integrates both interoceptive (motivational) and exteroceptive (visual) information. The MoHA model reproduces hippocampal firing correlates for different motivational states. We show that the representations of MoHA allow a synthetic agent to learn and deploy efficient trajectories in a foraging task, laying the foundation for self-regulated multipurpose reinforcement learning.

This chapter was adapted from: Rosado, O. G., Amil, A. F., Freire, I. T., Vinck, M., & Verschure, P. F. (2025). Motivational cognitive maps allow robot biomimetic autonomy. 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 14402–14408

4.1 Introduction

Navigation is a fundamental capability for autonomous agents, both biological and artificial. In robotics and artificial intelligence, navigation algorithms have primarily relied on exteroceptive sensory information, such as visual, lidar, or depth data, to construct maps and determine optimal paths in an environment. These approaches leverage techniques such as simultaneous localization and mapping (SLAM), reinforcement learning, and graph-based path planning to enable efficient movement through space. While these methodologies have demonstrated success in a variety of tasks, they remain fundamentally constrained by their reliance on external sensory cues and predefined objectives. The resulting navigation strategies optimize spatial efficiency but lack a key aspect of biological autonomy: the integration of internal states to dynamically modulate behavior (Mossio & Moreno, 2015).

Unlike traditional AI approaches, biological navigation is inherently goal-oriented and driven by an organism's internal needs (M. Wilson, 2002). Unlike artificial agents, where goals are typically defined in an explicit and static manner, goal selection in biological autonomy is dynamically influenced by the physiological state of an agent and changing environmental conditions (Hull, 1945). For instance, an animal's motivation to seek food or water depends on its current metabolic needs rather than a predefined objective. This fact highlights a crucial limitation in current AI-driven navigation: the absence of an adaptive goal-selection mechanism that integrates both external and internal information.

The mammalian hippocampus is widely recognized as a key brain structure involved in spatial cognition. Lesion studies have shown that hippocampal damage impairs performance in spatial tasks (Douglas, 1967), and the discovery of spatially tuned cells (O'Keefe & Dostrovsky, 1971) supports Tolman's theory of a cognitive map (Tolman, 1948). This cognitive map theory suggests that animals construct internal representations of their environment to guide behavior rather than simply relying on stimulus-response associations. O'Keefe's seminal work on place cells (O'Keefe, 1976) further substantiated this idea, demonstrating that hippocampal neurons encode specific spatial locations in an environment. Notably, this work also demonstrated that the firing activity of hippocampal neurons is not solely influenced by spatial cues, but that they exhibit rate remapping as a function of features of sensory cues and behavioral task variables (Bos et al., 2019; Leutgeb et

al., 2005), and that they can be sensitive to visual and tactile information [See (Aronov et al., 2017) for more recent evidence in the auditive domain].

Beyond spatial encoding, evidence suggests that the hippocampus integrates interoceptive information relevant to an organism's motivational state. Early studies reported the existence of approach/consummatory cells that differentially fire during eating and drinking behaviors (O'Keefe, 1976; Ranck Jr, 1973). Kennedy and Shapiro (Kennedy & Shapiro, 2009) identified motivational place cells in the hippocampus, which alter their activation patterns depending on whether an animal is food- or water-deprived. These findings suggest that hippocampal neurons do not merely encode the organism's spatial location but also track the organism's internal state, allowing for adaptive behavior selection based on physiological needs.

Based on homeostatic and allostatic principles (Bernard, 1865; Cannon, 1939; Sterling & Eyer, 1988), previous models have successfully replicated the hypothalamic orchestration of internal needs (Guerrero-Rosado et al., 2022; Sanchez-Fibla et al., 2010). Thus, allostatic models allow artificial agents to survive in dynamic environments by enabling goal selection based on their motivational states. However, these models do not explain how cognitive maps develop and support adaptive behaviors. Other models aiming to explain hippocampal functioning have successfully captured cognitive map features without explaining the influence of motivational states (Fibla et al., 2010; Freire et al., 2024a; Maffei et al., 2015). Homeostatic reinforcement learning models ground the reward value of stimuli in the internal state of the agent, facilitating more biologically realistic learning and decision-making processes (Keramati & Gutkin, 2014; Laurençon et al., 2021; Maffei et al., 2015). While these approaches contribute valuable insights, they still rely on predefined representations of cognitive maps and do not fully capture the adaptive, motivationally driven nature of hippocampal spatial encoding.

Our approach builds upon the idea that the hippocampus compresses sensory information into meaningful representations, a function well captured by biologically-inspired convolutional autoencoders (Amil et al., 2024; Benna & Fusi, 2021; Santos-Pata et al., 2021a, 2021b). Specifically, we introduce a motivational hippocampal autoencoder (MoHA) model that integrates interoceptive (motivational) and exteroceptive (visual) information to generate compressed state representations. This architecture enables an artificial agent to encode not only spa-

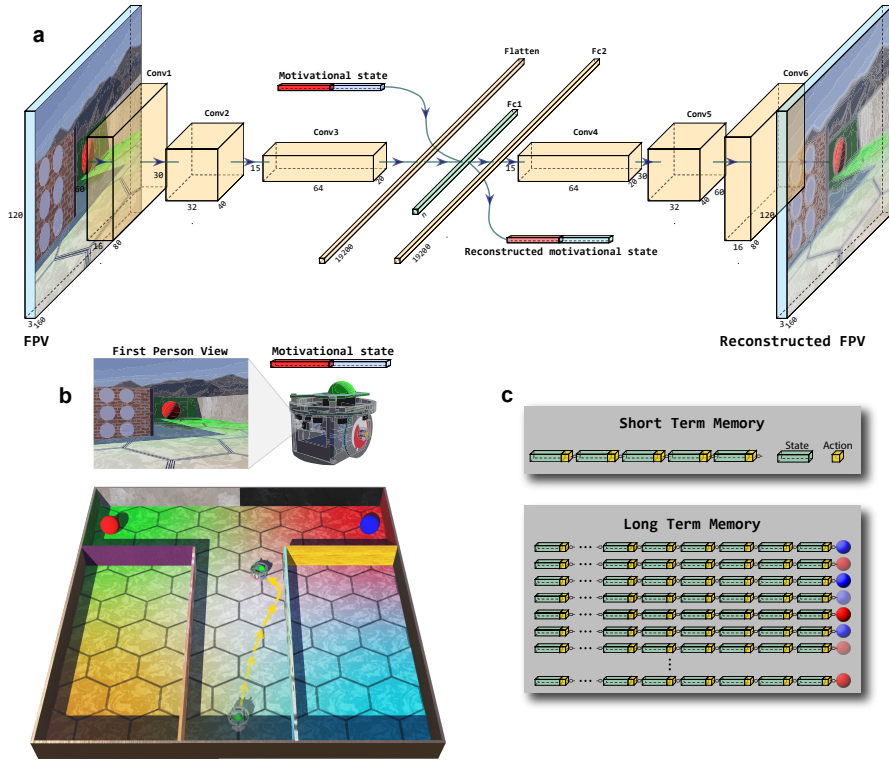


Figure 4.1. Motivational hippocampal autoencoder (MoHA). **a)** The MoHA uses the robot's First Person View (FPV) and its motivational state as input to build a joint compressed representation in its latent vector (green). Based on the latent vector, the decoding side of the network builds reconstructions of the inputs that, in turn, drive error backpropagation. **b)** Every trial of the spatial task starts with the agent at the beginning of the central alley and with a random dominant motivation. Through actions, the agent navigates the environment, sampling it until it finds a reward. At every step, its visual and motivational experience is passed through the MoHA to generate joint compressed representations of its state. **c)** These states, together with the actions that led to them, are stored in a short-term memory buffer as couplets. Every new couplet is linked to its predecessor, forming sequences that encode the current behavioral plan. If this plan results in reward gathering, the sequence stored in short-term memory moves to long-term memory together with the obtained reward value. Note that shorter trajectories result in a higher reward, as indicated by the color intensity of the reward stimuli.

tial locations but also how these locations relate to its internal needs. Moreover, the integration of Sequential Episodic Control (SEC) (Freire et al., 2024a) enhances the agent’s ability to store and retrieve sequential experiences, improving sample efficiency and decision-making by leveraging structured memory traces. The combination of these two elements, the MoHA model and SEC, offers a step toward achieving more flexible, self-regulated artificial intelligence that aligns with biological autonomy.

4.2 Methods

For motivational cognitive maps, we adopt the sparse autoencoder approach (Amil et al., 2024) and expand it to account for motivational information, in addition to the original visual input (Fig. 4.1a). In this manner, the objective function of our model aims to reduce the reconstruction error of both the visual and the motivational input:

$$\mathcal{L} = \frac{1}{m} \|\mathbf{X} - \hat{\mathbf{X}}\|_2^2 + \frac{1}{m} \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2 + \frac{\lambda}{mn} \|\mathbf{I}_n - \mathbf{Z}^\top \mathbf{Z}\|_F \quad (4.1)$$

where the first and second terms are the mean squared error (MSE) between inputs and their reconstructions, $\mathbf{X}$ denoting visual and $\mathbf{Y}$ motivational inputs, encouraging the latent vector $\mathbf{Z}$ to preserve input information. The second term is an orthonormal activity regularization term, whose strength is controlled by λ , pushing the Gramian $\mathbf{Z}^\top \mathbf{Z}$ towards the identity matrix $\mathbf{I}_n$. Finally, m denotes the batch size and n the number of neurons in $\mathbf{Z}$ [See (Amil et al., 2024) for a complete description of sparse autoencoders].

In the navigational task, Sequential Episodic Control (SEC) uses the autoencoder’s latent vectors as the agent’s state to create and optimize efficient behavioral plans. As the agent navigates the environment (Fig. 4.1b), every state is coupled with the action that led to that state, and a sequential representation of the current behavioral plan is created by linking these state-action couplets. Sequences are maintained in a short-term memory (STM) buffer. However, once the agent finds a rewarding stimulus, this sequence is stored in a long-term memory (LTM) system together with the corresponding reward value (Fig. 4.1c).

Through experience and familiarization with the environment, the agent starts experiencing states similar to those reward-related states

stored in LTM. Thus, it can transcend from random to memory-guided navigation by retrieving actions that were beneficial in the past. This environmentally mediated synergy between perception and behavior has been shown to systematically bias the agent’s input sampling, which in turn stabilizes behavior (Freire et al., 2024a; Maffei et al., 2015; P. F. Verschure et al., 2003). Both STM and LTM systems follow a First-In-First-Out rule once these memory systems reach an arbitrary storage limit. [See Supp. Material A for MDP formulation and (Freire et al., 2024a) for a complete description of Sequential Episodic Control].

Two distinct experiments were carried out with the aim of understanding the MoHA both as a standalone model and as a component of a cognitive architecture. First, we train the MoHA with images gathered during random exploration in a Webots (Webots, n.d.) simulated T-maze. Each of these images was randomly associated with one of two possible motivational states. To gain an understanding of the functioning of the MoHA, different model configurations were analyzed. Specifically, we considered the effects of environment visual complexity, embedding size (i.e., the number of units in the latent vector), and the motivational vector size. Second, a pre-trained MoHA was embedded in a simulated e-puck mobile robot that used SEC to learn appropriate goal-oriented behavioral plans. This agent started the experiment with empty STM and LTM systems and used spatial and motivational information to improve its navigation to the correct reward site, which was conditional on its current motivational state.

4.3 Results

Motivational Cognitive Maps

In order to identify motivationally tuned units in the latent vector of the MoHA, we computed a Gain Modulation Score (GMS) for each unit in the latent vector as:

$$\text{GMS}_i = \frac{|\max(fr_{i,m1}) - \max(fr_{i,m2})|}{\max(fr_{i,m1}) + \max(fr_{i,m2})} \quad (4.2)$$

where fr_i is the firing rate (i.e., activation) of unit i and m the motivational state for which the firing rate was computed.

The distribution of GMSs across units exhibited a trimodal pattern, suggesting the presence of distinct subpopulations with different modulation properties. Specifically, the histogram revealed peaks at both low and high gain modulation values, with relatively fewer units exhibiting intermediate modulation levels around $\text{GMS} \approx 0.5$. Notably, there was a pronounced accumulation of units at $\text{GMS} \approx 1$, indicating a substantial subset of units that were fully suppressed in one of the two conditions. 4.2a shows the distribution of GMSs for a reference model trained in the T-maze v2 with a latent vector of 400 units and a motivational input size of 10^3 .

Based on the observed distribution, we categorized units into three distinct groups according to their modulation characteristics (Fig. 4.2b):

- **Binary units** with $\text{GMS} > 0.99$, that is, active in one motivational context while almost completely inactive in the other.
- **Rate remapping units** with $0.5 \leq \text{GMS} \leq 0.99$, showing intermediate levels of gain modulation, reflecting changes in firing rate between contexts without complete suppression or activation.
- **Stable units:** with $\text{GMS} < 0.5$, exhibiting minimal modulation between contexts, such that their firing properties remain relatively stable.

This classification method differs from that used by Kennedy and Shapiro (Kennedy & Shapiro, 2009), which used firing-field correlation to quantify rate remapping. However, the unit class distribution of our reference model still resembles that found in physiological data (Fig. 4.2c) and, when comparing gain modulation and field correlation, low gain-modulated (stable) units showed high correlation values (Fig. 4.2d). Thus, our classification approach not only captures spatial but also rate remapping when encoding states under two different motivational contexts. Indeed, rate remapping has been reported under different conditions, such as the discrimination of different sensory cues (Leutgeb et al., 2005) and when analyzing social place cells (Bos et al., 2019).

To determine whether visual information affects unit class distribution in the MoHA, we trained the model under three spatially identical yet visually distinct T-maze configurations. Our results do not show an apparent alteration of the unit class distribution driven solely by visual differences in the environment (Fig. 4.2e) (Supp. Material B). However,

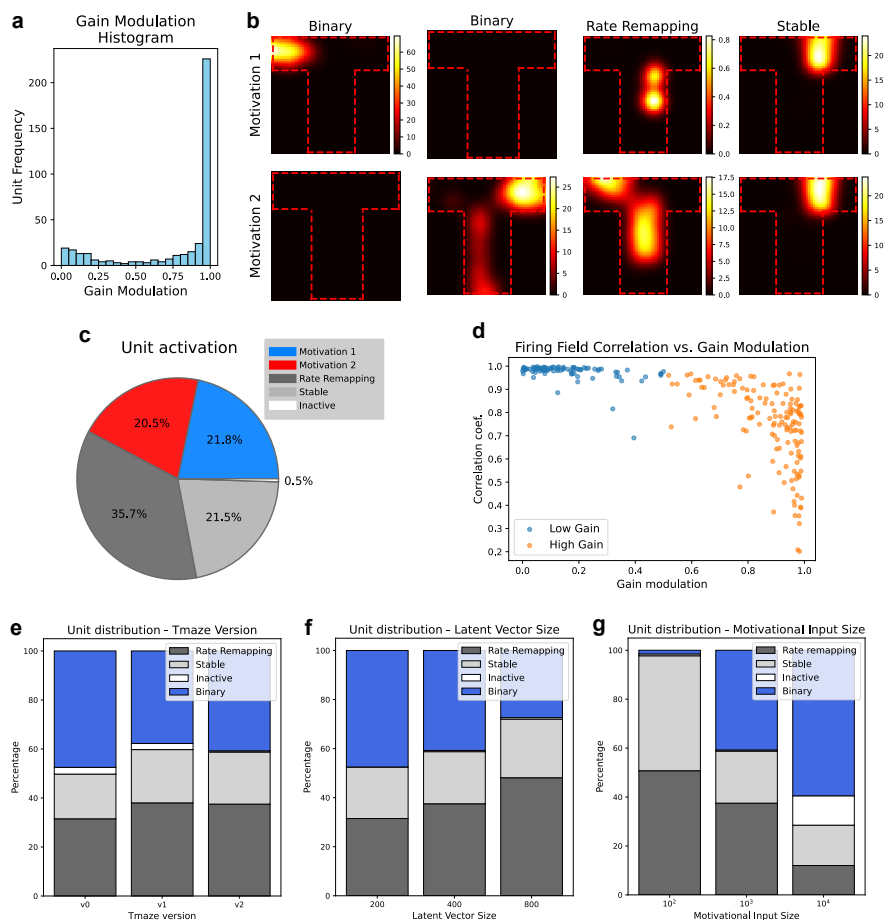


Figure 4.2. Motivational Cognitive Maps. **a)** Histogram of Gain Modulation Scores (GMSs) for a reference model trained in T-maze v2 with 400 units in its latent vector and motivational input size of 10^3 . **b)** Rate maps of example units during motivation 1 and 2 separately. Notice that binary units display activation for only one of the two motivational states. Rate remapping units may get active in a similar location but with a distinct firing rate. Firing fields of stable units remain unaffected by motivational states. **c)** Unit class distribution for the reference model. Binary units were sub-classified based on their preferential activation to one of the two motivational states. **d)** Firing field correlation is related to Gain Modulation Scores. As units display lower gain modulation, they become more correlated across motivational states. **e)** Unit distribution comparison across T-maze versions (Appx. C). **f)** Unit distribution comparison across latent vector size. **g)** Unit distribution comparison across motivational input size.

we did find that the unit class distribution was sensitive to the architectural features of the model. The percentage of binary units tended to decrease as we increased the size of the latent vector (Fig. 4.2f); an effect accompanied by an increase in the proportion of rate-remapping units. The opposite effect occurred when increasing the size of the motivational input (Fig. 4.2g).

A motivationally-guided SEC agent

To elucidate whether motivational cognitive maps generated by our model are adequate to guide spatial learning, we integrated the MoHA with SEC. A simulated mobile e-puck robot endowed this architecture to learn efficient strategies in a spatial decision task, where motivational information is crucial for selecting the correct goal. To compare performance, a control robot was endowed with a similar architecture and started every trial with a random dominant motivation, but it embodied a regular (i.e., non-motivational) hippocampal autoencoder. The experiment was divided into two parts: First, both control autoencoder and MoHA were pre-trained with images gathered during random exploration of the environment. As in the previous experiment, the MoHA paired each of these images with a random motivation. Second, SEC enabled learning in the spatial decision task. Here, both motivational and control models shared the same STM and LTM of size 50 and 200, respectively.

During the pretraining of the motivational autoencoder, visual reconstruction error remained lower for the control model, which is an expected result since the control autoencoder did not dedicate resources to the encoding of motivational information, as is the case with the MoHA (Fig. 4.3a). Despite the fact that differences in the reconstruction error were visually appreciable, both the control and motivational autoencoder models reconstructed the input images reasonably well (Fig. 4.3b). Moreover, when analyzing the unit's spatial information (W. Skaggs et al., 1992) (Supp. Material C), the control autoencoder did not differ from the motivational autoencoder during motivation 1 or 2 (Fig. 4.3c).

In the spatial decision-making task, the motivationally-guided agent successfully learned to navigate to the correct reward given the current motivation. After approximately a hundred trials, the agent had already navigated to the correct reward more than 80 percent of the time, and its probability of choosing the correct reward kept improv-

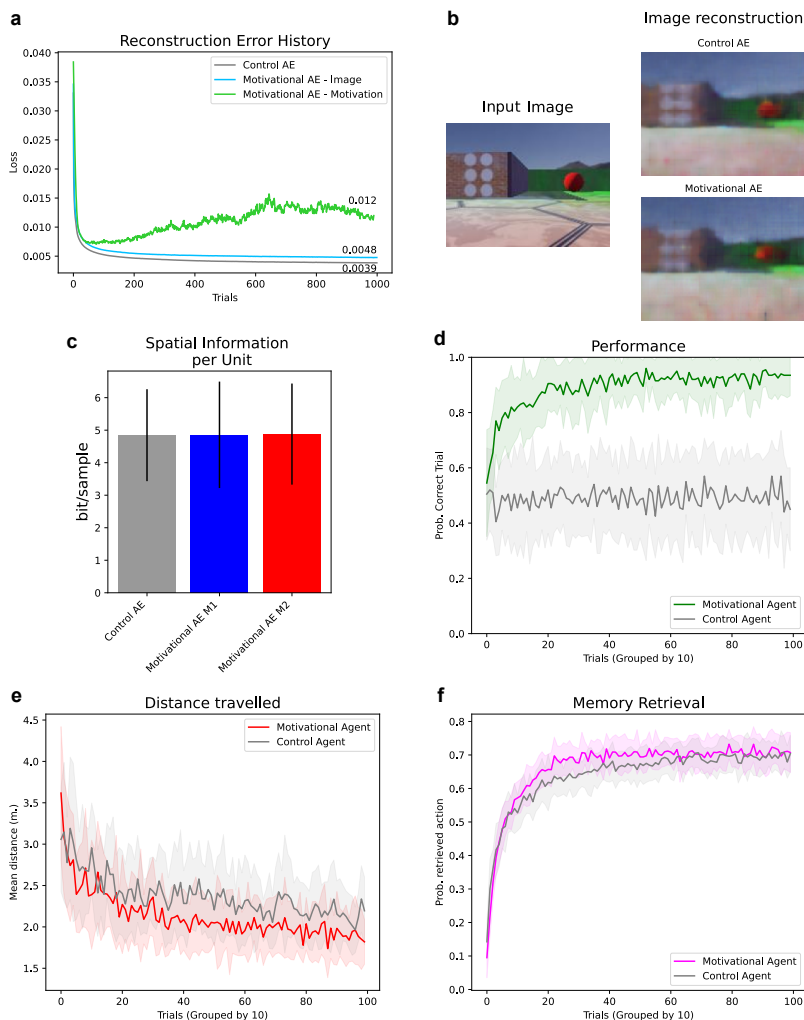


Figure 4.3. Performance of a motivationally-guided artificial agent in the Tmaze foraging task. **a)** Reconstruction error history for both control and MoHAs (Motivational Hippocampal Autoencoders). **b)** Image reconstruction example for both control and MoHAs. **c)** Spatial information comparison, split into motivations 1 and 2 for the MoHA. **d)** Agents' performance, measured as the probability of gathering the correct reward (given the trial's dominant motivation). **e)** Trajectory optimization across trials. **f)** Memory retrieval, measured as the probability of performing a memory-guided action over a random one. Plots d, e, and f inform about mean measures gathered during 20 experiments for both motivational and control agents.

ing (Fig. 4.3d). On the other hand, the control agent showed poor performance, with a probability of choosing the correct reward around chance levels during the entire experiment. Despite this difference in performance, both motivational and control agents learned to optimize their trajectories toward rewards (Fig. 4.3e) and increased the rate of memory-guided actions performed along the trial (Fig. 4.3f). These results demonstrate that although the control agent successfully learned to navigate toward reward sites, only the motivational agent could adapt its selected target location based on the currently dominant motivational context.

4.4 Discussion

In the present work, we developed a motivational hippocampal autoencoder (MoHA) that jointly encoded motivational and spatial states. This joint encoding led to the development of motivational-context-dependent place fields, exhibiting rate remapping.

Hippocampal rate remapping has been previously described based on changing sensory cues, or the presence of another agent in the same environment (Bos et al., 2019; Leutgeb et al., 2005). This phenomenon highlights the key distinction between global remapping, in which place fields shift to other locations, and rate remapping, in which firing rates are gain modulated (Leutgeb et al., 2005).

Our results strongly indicate that MoHA exhibits rate remapping, as its unit distributions showed gradual gain modulation rather than full shifts in firing locations when encoding different motivational states. A rate-based classification approach allowed us to identify binary, stable, and rate remapping units in a distribution similar to that found in physiology by Kennedy and Shapiro (Kennedy & Shapiro, 2009). Nonetheless, their classification approach was based on place-field correlations, an approach that cannot distinguish rate remapping from global remapping. Further experiments are needed to determine whether MoHA could also support global remapping. A potential future direction is to test whether the integration of a FiLM layer (Perez et al., 2018) within the MoHA drives global remapping while conserving spatially constrained firing fields.

In addition to exploring MoHA's neural representation, we have tested its suitability in motivationally guided foraging tasks, where multiple reward identities must be considered. Our integration of MoHA with Se-

quential Episodic Control (SEC) (Freire et al., 2024a) leveraged MoHA's state representations to generate optimal behavioral policies. This resulted in efficient performance in the T-maze foraging task, demonstrating that motivational cognitive maps are necessary for an agent to learn reward-directed strategies.

The interoceptive modulation of hippocampal activity remains an open question, with potential mechanisms involving the hippocampus's unique connectivity and receptor expression. The hippocampus interfaces with the lateral ventricle, providing access to internal markers such as hormones, metabolites, and temperature fluctuations. Additionally, hippocampal neurons express a diverse array of receptors, far more than the average neuron, allowing them to integrate multimodal sensory and interoceptive signals (Buzsáki & Tingley, 2023). This extensive receptor profile enables the hippocampus to link bodily states to cognitive processes. Understanding how these mechanisms contribute to the formation of motivational cognitive maps is crucial for advancing biologically inspired artificial intelligence.

Beyond its neuroscientific relevance, our findings have direct implications for robotic autonomy. Traditional AI-based navigation approaches rely on externally defined goals and operate under static reward structures, limiting their ability to dynamically adapt to changing internal states. By contrast, MoHA enables an agent to generate self-regulated cognitive maps, where goal selection is contingent on its motivational state rather than predefined objectives. This represents a significant step toward biologically inspired robotic autonomy, allowing for more adaptive and goal-directed decision-making.

MoHA potentially addresses two core challenges in robotics: First, it could enable robots to prioritize competing goals and separate strategy representations based on their internal states, partially solving task management and action orchestration problems. Second, by continuously updating cognitive maps based on both sensory and motivational inputs, MoHA provides a foundation for lifelong reinforcement learning in dynamic environments, where an agent can adapt its decision-making over time.

Future work should explore whether MoHA could encode a larger set of motivational states, opening the door to biologically-constrained multipurpose reinforcement learning. The scalability of MoHA to larger real-world scenarios also deserves future investigation.

Acknowledgements

This study was funded by Counterfactual Assessment and Valuation for Awareness Architecture—CAVAA (European Commission, EIC 101071178)

Code availability

The code for the Motivational Hippocampal Autoencoder (MoHA) is publicly available at the following GitLab repository:

https://gitlab.com/OscarGuerreroRosado/motivational_hippocampal_autoencoder

This repository includes both the offline MoHA training notebooks and the online Webots experimental setup.

Supplementary Material

A. Markov Decision Process (MDP) formulation

The underlying learning and decision-making process of the agent on the navigational task can be framed as a Markov Decision Process (MDP). $\langle S, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma \rangle$, where:

- S is the state space, where each state $z_t \in \mathbb{R}^n$ is a latent representation produced by the Motivational Hippocampal Autoencoder (MoHA). The encoding $z_t = \phi(x_t, m_t)$ compresses the agent’s first-person visual observation x_t , modulated by its internal motivational state m_t , into a single high-dimensional vector.
- $\mathcal{A}$ is the discrete action space. In our implementation, $\mathcal{A}$ includes forward locomotion actions with varying degrees of left or right turning.
- $\mathcal{T}$ is the environment’s transition function, implicitly defined by the physics of the simulation. While we do not model $\mathcal{T}$ explicitly, partial transition dynamics are captured through the sequential storage of state–action–reward trajectories in memory.
- $\mathcal{R}$ is the reward function, assigning a positive reward (e.g., $r = 1$) when the agent reaches a goal that matches its current mo-

tivational state m_t , and a small non-zero reward $r = 0.001$ otherwise. This baseline reward avoids numerical issues during reward-based memory retrieval in the SEC mechanism.

- $\gamma \in (0, 1]$ is the discount factor.

The policy $\pi(a_t|z_t)$ is implemented through *Sequential Episodic Control (SEC)*, a non-parametric, memory-based controller that retrieves temporally ordered sequences of past experience (Freire et al., 2024a). SEC selects actions based on perceptual similarity in latent space and motivational alignment with stored episodes, enabling goal-directed behavior through memory reuse.

B. T-maze versions

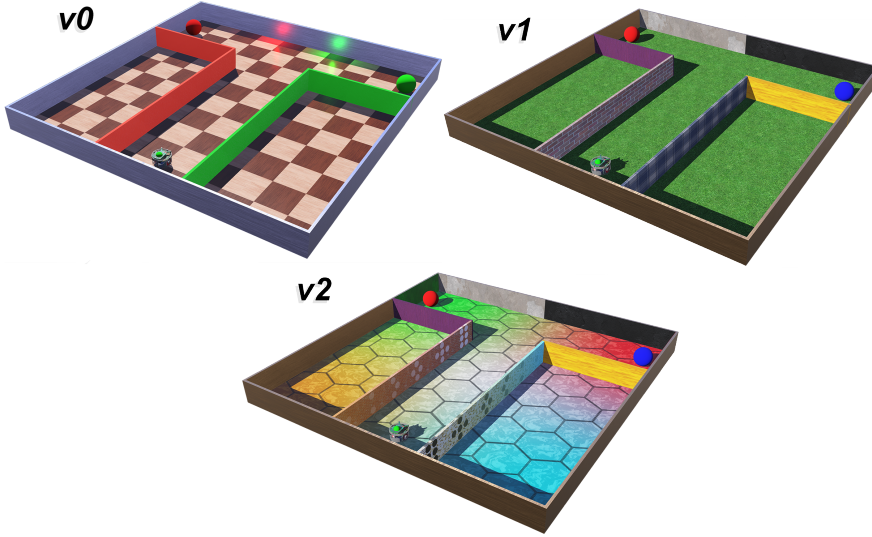


Figure 4.4. T-maze versions. Our Motivational Hippocampal Autoencoder was trained on three versions of the T-maze that differed only in their visual appearance (textures and colors of walls and floor). The spatial layout, reward locations, and task structure were identical across all versions, ensuring that only exteroceptive visual cues were varied while environmental dynamics remained controlled.

C. Spatial information

Spatial information is computed as:

$$Information = \sum_{i=1}^N p_i \frac{\lambda_i}{\lambda} \log_2 \frac{\lambda_i}{\lambda},$$

where $i = 1, \dots, N$ denotes nonoverlapping spatial bins dividing the environment, p_i is the occupancy probability of bin i given the agent's random sampling distribution, λ_i is the mean unit activation for bin i , and λ is the overall mean activation of the unit in the environment.

5

CHAPTER 5

Socially Adaptive Cognitive Architecture for HRC in Industrial Settings

ABSTRACT

This chapter introduces DAC-HRC, a novel cognitive architecture designed to optimize human-robot collaboration (HRC) in industrial settings, particularly within the context of Industry 4.0. The architecture is grounded in the Distributed Adaptive Control theory and the principles of joint intentionality and interdependence, which are key to effective HRC. Joint intentionality refers to the shared goals and mutual understanding between a human and a robot, while interdependence emphasizes the reliance on each other's capabilities to complete tasks. DAC-HRC is applied to a hybrid recycling plant for the disassembly and recycling of Waste Electrical and Electronic Equipment (WEEE) devices. The architecture incorporates several cognitive modules operating at different timescales and abstraction levels, fostering adaptive collaboration that is personalized to each human user. The effectiveness of DAC-HRC is demonstrated through several pilot studies, showcasing functionalities such as turn-taking interaction, personalized error-handling mechanisms, adaptive safety measures, and gesture-based communication. These features enhance human-robot collaboration in the recycling plant by promoting real-time adaptation between humans and robots. The DAC-HRC architecture aims to contribute to the evolution of HRC paradigms by paving the way for more seamless and efficient collaboration in Industry 4.0 by relying on socially adept cognitive architectures.

This chapter was adapted from: Freire, I. T., Guerrero-Rosado, O., Amil, A. F., & Verschure, P. F. (2024b). Socially adaptive cognitive architecture for human-robot collaboration in industrial settings. *Frontiers in Robotics and AI*, 11, 1248646

5.1 Introduction

The increasing automation in the industry over the past decades has been partially driven by the adoption of robots, which have proven to be valuable tools for handling heavy, risky, and repetitive tasks. By automating these tasks, robots have helped alleviate the burden on human workers, contributing to improved safety, efficiency, and productivity (Cette et al., 2021; Mazachek, 2020). However, robotic solutions have their own limitations; they tend to have restricted operational capacity in terms of degrees of freedom and decision-making and therefore they do not perform well outside of highly controlled and structured environments (Vysocky & Novak, 2016). As a result, complete automation might be neither feasible nor desirable (Charalambous et al., 2015; Weiss et al., 2021), a discussion further amplified by recent advances in AI (P. Verschure et al., 2020).

Instead, the future of Industry 4.0 (Lasi et al., 2014) lies in the collaboration between humans and robots, capitalizing on the strengths of both in a manner that is both beneficial for the health and well-being of the workforce and productive for companies. Human-Robot Collaboration (HRC) is poised to become a key component of Industry 4.0 (Baratta et al., 2023), with the primary goal of creating a safe environment for humans and robots to collaborate effectively. This transition from traditional automation to Industry 4.0 will be marked by a transformation involving the use of the latest advancements in information and communication technologies (ICTs) (Robla-Gómez et al., 2017).

However, the transition to HRC in industrial settings faces several technical and scientific challenges (Inkulu et al., 2021). Physically close collaboration between human workers and industrial robots has been limited so far, primarily due to safety concerns, such as potential collisions causing injury to human operators (Robla-Gómez et al., 2017; Vysocky & Novak, 2016). Recent advances in collaborative robotics, including the emergence of cobots, now allow for safer and closer interactions between humans and robots (Weiss et al., 2021). These advances in collaborative technologies call for novel paradigms to design collaboration in hybrid industrial settings that are in line with the ambitions of Industry 4.0.

One prime example of such a setting has been the implementation of a hybrid human-robot system in a recycling plant for Waste Electrical and Electronic Equipment (WEEE) products under the umbrella of the EU-funded HR-Recycler project (Axenopoulos et al., 2019). This type of

environment poses a unique set of challenges for human-robot interaction (HRI) and collaboration (HRC). The noisy, dynamic, and rapidly changing conditions of a recycling plant lead to safety concerns when humans and robots must work together in such settings (Robla-Gómez et al., 2017; Vysocky & Novak, 2016). Moreover, the specific tasks involved in the recycling process demand physically close collaboration and interaction between workers and robots.

Collaborative robots, or cobots, have been developed within the last few decades across several domains, such as social and assistive robots (Rosenberg-Kima et al., 2019; Sandoval et al., 2018). These cobots, however, serve different purposes than the industrial robots used for manufacturing. As such, a new approach to human-robot collaboration, specifically tailored for the unique requirements of a dynamic industrial context, is necessary (Castro et al., 2021).

In the context of Industry 4.0, human-robot collaboration (HRC) takes on a new dimension. It becomes a complex sociotechnical system where agency—the capacity to act—is not solely attributed to humans. Instead, it is shared among humans and non-human agents, such as machines, robots, sensors, and software (Weiss et al., 2021). This shift in understanding is crucial as it acknowledges the increasingly interactive roles between humans and cobots in an Industry 4.0 environment. It also highlights the emergence of new constellations of shared control and distributed agency, which are key aspects of this new industrial paradigm.

In order to address the challenges of incorporating cobots in hybrid industrial settings, in this chapter we present a novel systems-level control paradigm for the design and implementation of cognitive architectures for HRC. To this end, we introduce DAC-HRC, a novel cognitive architecture designed for human-robot collaboration in industrial settings. In the next sections of the introduction, first, we outline the key principles for human-robot collaboration upon which the cognitive architecture is based, with special emphasis on the notions of joint intentionality and interdependence. We then introduce the Distributed Adaptive Control perspective for building HRC and highlight how each of the specialized modules of the DAC-HRC cognitive architecture is related to these principles and the state-of-the-art of each implemented functionality.

Principles for Human-Robot Collaboration in Industrial Settings

The advent of Industry 4.0 has ushered in a new era of Human-Robot Collaboration (HRC) in industrial settings, necessitating a reevaluation of interaction paradigms to accommodate the dynamic and intricate dance between human workers and cobots. The HR-Recycler challenge illustrates the complexity of such interactions within the noisy and dynamic environment of a recycling plant (Axenopoulos et al., 2019). In this section, we enumerate the most relevant principles for effective Human-Robot Collaboration considering the HR-Recycler's context of implementation, while conciliating the perspective of different authors that have explored this in the past (Adams, 2005; Bauer et al., 2008; Goodrich & Olsen, 2003; Levinson, 2020).

- **Implicit Switch Modes:** The system must fluidly alternate between various interaction modes, adapting to the human worker's context without burdening them (Bauer et al., 2008).
- **Natural Cues:** Intuitive interaction is facilitated by leveraging humans' inherent understanding of natural signals, enabling humans to communicate with robots using familiar gestures and symbols (Goodrich & Olsen, 2003).
- **Direct World Manipulation:** Interactions are designed to serve the ultimate purpose of task completion in a tangible world, allowing humans to directly influence robotic behavior to navigate the unpredictable physical environment of industrial settings (Adams, 2005).
- **Information Manipulation:** Information presented by the robot must be actionable, supporting the human worker's decision-making processes and promoting goal-oriented collaboration (Goodrich & Olsen, 2003).
- **Attention Management:** The design of HRC interactions should cater to the cognitive limitations of human attention, ensuring that critical information is highlighted and that potential attentional lapses are mitigated (Adams, 2005).
- **Situational Awareness:** Maintaining an acute awareness of the robot's internal and external state is paramount, enabling human workers to anticipate robotic actions and intervene when necessary (Goodrich & Olsen, 2003).
- **Safety:** Paramount to any HRC system is the unwavering commitment to human safety, ensuring that robots can navigate the

potential hazards of industrial tasks without endangering human collaborators (Goodrich & Olsen, 2003).

5.1.0.1 Joint intentionality and interdependence as core principles for industrial HRC

Beyond the general HRI principles described above, the DAC-HRC architecture incorporates two core principles coming from our current understanding of the origins of human collaboration: interdependence and joint intentionality.

Joint intentionality refers to the shared mental states and cooperative activities that arise when individuals engage in collaborative endeavors. Research on social cognition posits that shared intentionality is a unique feature of human cognition, setting us apart from other primates (Tomasello et al., 2005). It manifests in the form of shared goals, joint attention, and mutual knowledge among individuals working together. For instance, when two people collaborate to lift a heavy object, they share a common goal (i.e., moving the object) and are aware of each other's intentions, roles, and actions.

This concept is particularly relevant in the context of Human-Robot Collaboration (HRC), as it emphasizes the importance of mutual understanding, communication, and coordination between human workers and their robotic partners. In industrial HRC, developing systems capable of exhibiting joint intentionality is essential for ensuring more efficient and safer interactions between human workers and robots. In this context, developing shared intentionality in artificial and hybrid collaborative systems would imply the ability to (1) detect and predict human intentions, actions, and goals, (2) communicate its intentions, actions, and goals to human workers and (3) coordinate and adapt its behavior based on the shared goals and the feedback from the ongoing collaboration.

Interdependence is another foundational aspect of current theories of the evolution of human cooperation (Tomasello et al., 2012). It refers to the reliance of individuals on one another to achieve shared goals or complete tasks, and its key role in vital tasks for early humans such as obligate collaborative foraging (O'Madagain & Tomasello, 2022; Tomasello, 2009). Applied to the context of HRC, interdependence implies that both human workers and robots depend on each other's actions, skills, and knowledge to execute tasks effectively. The Interdependence Hypothesis (Tomasello et al., 2012) suggests that inter-

dependence fosters cooperation, as it encourages individuals to align their goals, share information, and coordinate their actions.

In HRC, task interdependence between humans and robots can motivate the design of systems that (1) recognize the skills and capabilities of human workers and adapt their behavior accordingly, (2) share task-related information with human workers, facilitating mutual understanding and efficient task execution (3) respond to changes in the task or environment, adjusting their actions to maintain effective collaboration.

To illustrate, consider a collaborative robot in an automotive assembly line. The cobot could be designed to recognize the specific skills of the human worker, such as their proficiency in installing certain parts. Based on this recognition, the cobot could adapt its behavior to complement the worker's skills, perhaps by preparing the necessary parts or tools for the worker's next task. Moreover, the cobot could share task-related information with the worker, such as the sequence of assembly steps or the status of the parts supply, facilitating mutual understanding and efficient task execution. Finally, the cobot should be able to respond to changes in the task or environment. For instance, if a part is missing or defective, the cobot could adjust its actions, perhaps by fetching a replacement part or alerting the worker to the issue.

In understanding and implementing these core principles of joint intentionality and interdependence, it becomes apparent that a sophisticated cognitive architecture is required: one that not only comprehends human social behaviors but also adapts and responds to the dynamic nuances of industrial settings. This necessity brings us to the Distributed Adaptive Control (DAC) approach, which provides a validated and biologically-inspired framework. The DAC approach, with its layered control system and emphasis on adaptability, is ideally suited to embed these principles into the fabric of human-robot collaboration. As we transition to exploring the DAC-HRC architecture, we will see how each of its specialized modules is designed to operationalize the principles of joint intentionality and interdependence, thereby creating a harmonious and effective collaborative environment between humans and robots in industrial settings.

The Distributed Adaptive Control Approach to Human-Robot Collaboration

A cognitive architecture is a modular control system that governs a robot's decision-making, information processing, and environmental interaction (Moulin-Frier et al., 2017; Vernon, 2022). The Distributed Adaptive Control (DAC) theory of mind and brain (P. F. Verschure, 2012) offers a robust theoretical foundation for such architectures, as it has been previously shown in various HRI scenarios (Fischer et al., 2018; Lallée et al., 2015; Moulin-Frier et al., 2017). DAC views the brain as a hierarchical system with multiple control layers, each crucial for adaptive behavior in diverse physical and social contexts (P. F. Verschure, 2012; P. F. Verschure et al., 2014). This biologically grounded modular modeling approach is especially suitable for addressing the HR-Recycler environment's challenges, which demand adaptive and goal-oriented actions.

Informed by the DAC framework, the DAC-HRC architecture we introduce in this chapter integrates four specialized modules that reflect key principles for effective HRC. Each module is tailored to specific principles, forming a cohesive and operational control system:

- **Task Planner:** Coordinates the proper disassembly steps for each device, organizes the disassembly procedure and the turn-taking between human and robot actions, centralizes task-related information among the DAC-HRC modules, and implements safe and robust error-handling protocols. The Task Planner reflects the principles of shared intentionality and interdependence, as it involves a mutual understanding between the human and robot about the sequence of tasks and reliance on each other's capabilities to complete these tasks. Moreover, by orchestrating robot control and human-robot interaction, it also embodies the principles of 'implicit switch modes' and 'direct world manipulation'.
- **Interaction Manager:** Serves as a multimodal, non-verbal communication interface, facilitating efficient communication and interaction between humans and robots. To achieve this, the module integrates multimodal channels of communication, ranging from audiovisual interfaces such as tablets to embodied gesture-based, communication. By handling natural embodied human-robot interaction based on gestures and adapting to the context of the information visualized on tablet devices, this module implements the principles of 'natural cues', 'attention management', 'in-

formation manipulation', and 'situational awareness'. By jointly visualizing in the tablet device the progress and information about the status of the human worker and the robot, this module also creates a sense of shared intentionality.

- **Moral Engine:** Acts as a context-aware adaptive safety mechanism, controlling the safety distances between humans and robots as well as the speed of the interactions, adapting them to the context and the preference of the human co-worker. It deals with the integration of the relevant environmental, social, and material information that comes from other modules to adapt the safety mechanisms of the human-robot collaboration, directly addressing the principles of 'safety' and 'situational awareness'. It dynamically adjusts robot behavior to align the safety measures with human preferences and the task context, also emphasizing the principle of interdependence.
- **Worker Model:** Creates an internal model of human workers, focusing on the principles of 'information manipulation', 'implicit switch modes', 'situational awareness', and 'interdependence'. This module handles information about the human worker, using it to adapt the robot's behavior in alignment with the worker's preferences. This module is instrumental in adapting the overall collaboration schemes to the human worker, enabling the robot to adjust its actions and fostering a collaborative relationship where both parties rely on and benefit from each other's strengths.

By integrating these modules within a single control architecture, DAC-HRC, we create a robust control system for HRC in industrial settings. This system not only adapts to different workers and scenarios but also fosters a mutual understanding and effective collaboration between human workers and robots, essential for the complex tasks in the HR-Recycler's recycling plant. The DAC-HRC cognitive architecture is composed of several specialized modules, and therefore each of them follows and implements a different function based on distinct state-of-the-art solutions from the literature.

5.1.0.2 Task Planner as a Hierarchical Finite State Machine

Task planners play a pivotal role in robotics, especially in enabling robots to adeptly navigate complex and unpredictable environments. In domains like electronic waste recycling operations, the ability of robots to perform a range of tasks, from sorting to processing diverse

types of devices and components, hinges on sophisticated task planning mechanisms (Alami et al., 2005). The cornerstone of contemporary task planning in robotics is the use of finite state machines (FSMs), revered for their simplicity and intuitiveness in modeling robot behavior amidst uncertainty (Foukarakis et al., 2014).

FSMs are essentially mathematical constructs encompassing a finite set of states, transitions between these states, and corresponding input/output events. This structure empowers robots with the ability to efficiently adapt their behavior in response to varying conditions, a feature crucial in the fluctuating environment of a recycling plant. The inherent simplicity of FSMs, however, can be a limitation when dealing with more complex behaviors.

To address this complexity, hierarchical FSMs (HFSMs) have emerged as a potent solution. HFSMs represent behaviors in a layered structure of FSMs, where each level corresponds to a specific subtask or behavior component (Johannsmeyer & Haddadin, 2016). This hierarchical arrangement facilitates a modular and scalable approach to task planning. By breaking down overall robot behavior into manageable subtasks, HFSMs offer a tailored solution to the multifaceted tasks encountered in electronic waste recycling. This approach not only enhances the robot's efficiency and adaptability but also allows for easier integration and updates to the task planning system as recycling requirements evolve.

Moreover, the incorporation of human-in-the-loop methodologies in task planning signifies a significant evolution in robotic systems. This approach involves integrating human feedback and inputs directly into the robot's control mechanism, enabling a more dynamic and adaptable interaction between humans and robots (Raessa et al., 2020). In the context of electronic waste recycling, this means that robots can be more responsive to human operator's preferences and needs, thereby enhancing collaboration efficiency and safety.

In implementing HFSMs, the Task Planner module within the DAC-HRC cognitive architecture embodies these principles, leveraging the hierarchical structure to manage complex tasks while remaining adaptable to the diverse challenges presented in electronic waste recycling. The module's design allows for seamless incorporation of human inputs, ensuring that the robotic system is not only responsive but also attuned to the nuanced needs of human workers. This integration of advanced FSMs within the DAC-HRC architecture illustrates a commitment to developing robotic systems that are both technically proficient

and collaboratively effective in complex industrial settings.

5.1.0.3 Interaction Manager as a multimodal non-verbal communication protocol

The DAC-HRC architecture's Interaction Manager advances the paradigm of multi-modal non-verbal communication, pivotal for intuitive and effective human-robot collaboration in industrial settings. In the human-centered HRI paradigm, an essential aspect of implementing a successful and effective HRI is building a natural and intuitive interaction (X. Wang et al., 2022). In recognition of the importance of non-verbal communication modalities, particularly in noisy industrial settings, the Interaction Manager eschews auditory channels in favor of gesture-based and tablet-based interfaces.

Gestures serve as a fundamental form of human communication, making them ideal for conveying rapid commands in human-robot interaction (HRI) (Pezzulo et al., 2019; Vouloutsi et al., 2020). Gesture-based communication harnesses the natural propensity for humans to use physical gestures, thereby facilitating a more immediate and universal form of interaction (H. Liu & Wang, 2018; Peral et al., 2022; X. Wang et al., 2022). The Interaction Manager incorporates a repertoire of shape-constrained gestures (Alonso-Mora et al., 2015) tailored to the communication needs specific to the HR-Recycler's project, which facilitates natural and intuitive interactions without extensive training (Vouloutsi et al., 2020) while also ensuring accurate recognition and interpretation by the robotic agents (Peral et al., 2022).

To complement gesture-based interactions and cater to scenarios necessitating more detailed information exchange, the architecture also integrates tablet-based communication. This method leverages interactive applications, which, while commonly used for teleoperation (Best & Moghadam, 2014; Luz et al., 2019; Yepes et al., 2013), are repurposed in the DAC-HRC to enhance the human-robot bond and situational awareness (Goodrich & Olsen, 2003). The tablet application provides a direct interface for the human worker to receive updates on the robot's internal state and environmental interpretations, aligning with key HRI principles (Adams, 2005; Goodrich & Olsen, 2003).

The combined use of gesture and tablet-based interactions by the Interaction Manager represents a state-of-the-art approach to non-verbal HRI. It successfully navigates the challenges of noisy industrial settings, where traditional verbal communication is untenable, and es-

establishes a robust, adaptive, and user-friendly communication system conducive to the dynamic requirements of the HR-Recycler's operations.

5.1.0.4 Safety and Moral Engine as an allostatic control system

A robot must not endanger a human under any circumstances. This premise, already formulated by Isaac Asimov in his famous "Three Laws of Robotics", is crucial for any robotic installation but especially for those promoting interaction and collaboration between humans and synthetic agents (Asimov, 2004). Importantly, in industrial settings interactions must be designed bearing in mind that robots occasionally will represent a source of danger to humans because of the tools they employ to perform hazardous tasks such as cutting, hammering, or moving heavy objects. Safety measures need to be implemented both according to the work to be performed and the human demands (Van Wynsberghe, 2020).

In the context of the WEEE recycling factory, humans and robots have to interact with a variety of different objects and tools and realize many changing sequences of actions in order to successfully complete their tasks (Axenopoulos et al., 2019). Although each of the robotic components has its own built-in safety mechanisms and corresponding certified ISO safety measures, the interaction of all these elements together will require an additional layer of control that can adapt their behavior to the requirements of the hybrid recycling plant while following human-centered design principles. This layer of control is the Moral Engine.

The Moral Engine within the DAC-HRC architecture goes beyond mere harm avoidance to actively promote cooperation (Freire et al., 2020c). By applying a synthetic and functional approach to morality, the Moral Engine not only adheres to the basic principles of safety but also engages in a deeper, more nuanced understanding of human-worker preferences, thus fostering a more cooperative and harmonious human-robot interaction. It also reflects the principles of shared intentionality and interdependence, as it involves a mutual understanding between the human and robot about the safety measures and reliance on each other's capabilities to maintain safety during the disassembly process.

The goal of the Moral Engine is to promote human-robot cooperation by building safer, more trustworthy, and personalized interactions

with human users (Christoforakos et al., 2021; Kok & Soh, 2020). It does so by regulating and adapting the robot's behavior to the particular human preferences of every user (Edmonds et al., 2019; Senft et al., 2019). In this way, it also serves as an extra layer of security for the system by integrating contextual information from the environment and using it to prevent potentially harmful situations (Yang et al., 2018). At the heart of this approach is the implementation of an allostatic control system (Guerrero-Rosado et al., 2022; Sanchez-Fibla et al., 2010). This system aims to ensure harm avoidance and promote cooperative behaviors, which are two fundamental aspects of moral behavior (Freire et al., 2020c).

In essence, the Moral Engine encapsulates the ethos of the DAC-HRC architecture —prioritizing human safety and introducing dynamic adaptability, thereby exemplifying a model of responsible and responsive artificial intelligence in industrial settings.

5.1.0.5 Worker Model based on human-centered design principles

User modeling systems rely on data gathering to create user models, either explicitly or implicitly (Luna-Reyes & Andersen, 2003). The advent of machine learning has driven many recommendation and adaptation applications for user modeling towards data-driven approaches (Kontogianni et al., 2018). One emerging technique being implemented in these data-driven user modeling practices is the Digital Twin concept, which generates or collects digital data representing a physical entity, emphasizing the connection between the physical and virtual counterpart through real-time information flow (Bruynseels et al., 2018; Negri et al., 2017).

Digital Twin technologies have been applied in various contexts, such as healthcare (Croatti et al., 2020) and human-robot interaction (Malik & Brem, 2021; Q. Wang et al., 2020; Wilhelm et al., 2021). In industrial settings, Digital Twins have been utilized for tasks like interactive welding, bridging human users and robots through bidirectional information flow, and benefiting novice welder training (Jokinen & Wilcock, 2015; Q. Wang et al., 2020). However, these data-driven approaches raise concerns regarding big data management, privacy, and trustworthiness, especially when applied to sensitive fields (Kumar et al., 2020). The Human-Centered AI paradigm aims to address these concerns by prioritizing methodologies that meet user needs while operating transparently, delivering equitable outcomes, and respecting privacy (Riedl, 2019; Xu, 2019). This approach also aligns with legislation such as

the European General Data Protection Regulation (GDPR) (Kloza et al., 2019).

The Worker Model module of the DAC-HRC cognitive architecture follows such human-centered design principles by maximizing functionality while minimizing the amount of data gathered from the user (Riedl, 2019; Xu, 2019). This design strategy ensures that the Worker Model respects user privacy while still providing effective support for human-robot collaboration in the disassembly of WEEE devices. The main goal of the Worker Model is to collect, process, and store all the relevant information regarding each user of the system and integrate it into one single, coherent data structure. This information is used by the DAC-HRC architecture to flexibly adapt the human-robot collaboration paradigm to the human partner. In other words, the Worker Model creates a virtualization of the human worker that allows the collaborative architecture to dynamically adjust its parameters to ensure a personalized interaction.

In essence, the Worker Model's integration into the DAC-HRC architecture not only enhances the adaptability of the human-robot collaboration paradigm but also embodies a human-centric focus in the design of these new technologies (Riedl, 2019; Xu, 2019).

The rest of the chapter is organized as follows: In the following methods section, we first describe the aCell, a specific experimental setup designed for the collaborative disassembly of WEEE devices. We then continue describing in detail each of the components of the DAC-HRC architecture along with its interactions. We proceed with a report of the main results showcasing the functionalities of the architecture across the different tested use cases, and conclude with a discussion of the main outcomes of the study, its limitations, implications, and future work.

5.2 Methods

The aCell experimental setup

The experimental setup consists of a specific spatial and technical configuration of an adaptive Collaborative Cell (aCell) designed for the collaborative disassembly of Waste Electrical and Electronic Equipment (WEEE) devices. The concept of an aCell represents an evolution in the

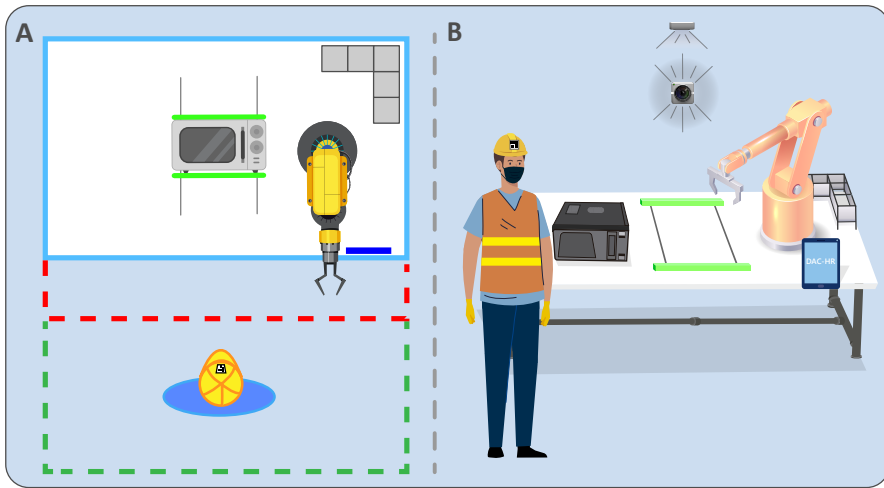


Figure 5.1. Experimental Set-up. **a)** Cenital view of the aCell. Green dashed lines illustrate the human working area, being limited by a safety threshold (red dashed lines). **b)** Complete configuration of the aCell including the human worker, the cobot, the WEEE device to be disassembled, the clamping tool, the tablet device where the Interaction Manager app is displayed, the cobot tool rack, and the cameras monitoring the behavior of both the human worker and robot.

way we approach task allocation in HRC (Axenopoulos et al., 2019). Traditional industrial HRI methodologies often focus on individual tasks within a single work cell, with the human and robot working in isolation on specific tasks. However, the aCell concept promotes a more holistic view of HRC that takes into account the interdependence between humans and robots. It envisions a dynamic, integrated system where the human and robot work together across multiple tasks, adapting to changes in the work environment and each other's capabilities. This approach aims to enhance the overall efficiency and effectiveness of the collaboration, rather than optimizing individual tasks in isolation.

An aCell is a dynamic and adaptive component of a hybrid factory, responsible for a specific task and for a given time period. The responsibilities, resource allocation, and overall positioning of its elements within the factory are dynamically assigned and adapted in real time with respect to the overall factory workflow demands, available skills, and available resources. In the context of our study, the cell consists of a human worker collaborating with a cobot, with each of them possessing specific, known skills. They operate as part of a joint intentional team with shared goals: to disassemble a series of Waste Electrical and Electronic Equipment (WEEE) devices.

The design of the aCell is grounded in the interdependence and joint intentionality between the human worker and the cobot. The components of the aCell are interdependent since effective task completion requires the combination of both human and robot capacities while sharing the same goals for disassembly. By taking into account the complementary skills and shared goals of the human-cobot dyad, the aCell can be seen as a single collaborative unit whose control is distributed. The DAC-HRC architecture we present in this chapter is designed as a control system to deal with such hybrid collaborative entities, by orchestrating the disassembly process while also taking into account human workers' safety, and promoting context and real-time adaptation in the dynamic and complex environment of the WEEE recycling plant.

In this work, the aCell is composed of two primary regions (Fig. 5.1): the open space, where the human worker performs tasks without hindrance, and the workbench, where the DAC-HRC synthetic actuators are strategically located.

A camera, oriented towards the open space, captures and processes information related to the human worker, such as their identity, position, and behaviors like gestures. On the workbench, WEEE materials are positioned for collaborative disassembly by the human worker and a COMAU Racer-5 collaborative robot (cobot). To further augment the functionality of the cobot, a tool rack is in place to house and arrange Racer-5 tools that are not currently in use. These tools include a screwdriver, a vacuum gripper, a finger gripper, and a cutting device.

Furthermore, to enable the cobot to gather information on the WEEE device, such as the condition of its components, an additional camera is directed toward the disassembly area. A mechanical clamping tool is also integrated with the workbench to stabilize the WEEE device while either the human worker or cobot performs actions on it.

Lastly, the workbench, and by extension, the aCell, are supplemented by a tablet display that enables a bilateral communication channel between DAC-HRC and the human worker, displaying relevant information (e.g., current task status), and serving as a medium for human feedback.

The Distributed Adaptive Control architecture for Human-Robot Collaboration

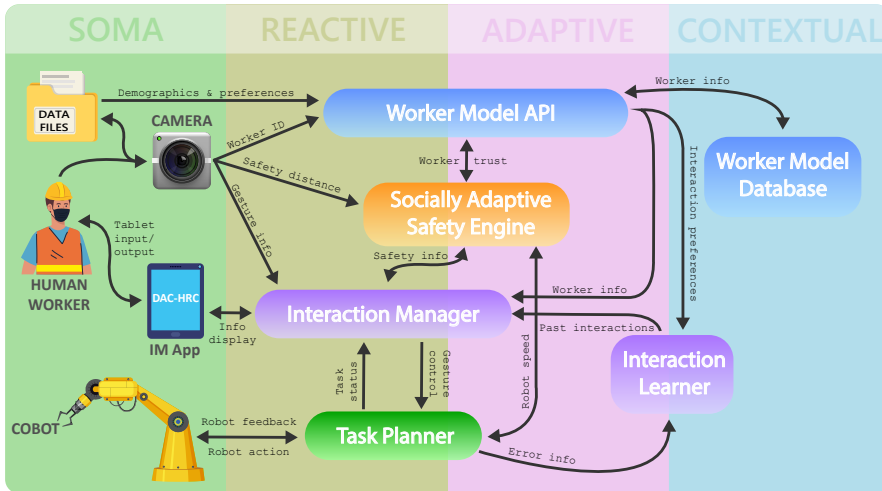


Figure 5.2. The DAC-HRC architecture for human-robot collaboration in industrial settings. DAC-HRC is structured in four layers of control (from left to right): Soma, Reactive, Adaptive, and Contextual; and composed of four specialized cognitive modules: the Task Planner, the Interaction Manager, the Moral Engine, and the Worker Model. The cognitive modules are distributed across various levels of the layered architecture, aligning with their control timescales and levels of information and sensory abstraction. The Soma Layer represents the physical components of the aCell, including the cameras, tablet, cobot, and human worker. The Reactive Layer houses the Interaction Manager and Task Planner modules, responsible for orchestrating human-machine communication and managing task allocation, respectively. The Moral Engine and the API component of the Worker Model, which adapts safety measures and processes real-time worker information, span both the Reactive and Adaptive Layers. The Contextual Layer is home to the Worker Model Database, storing long-term memory of past interactions and user preferences, and the Interaction Learner, which uses this information to adapt tablet display options based on past interactions.

The DAC-HRC cognitive architecture

The aim of the DAC-HRC architecture is to develop a robust human-robot collaboration control system for industrial settings that adapts to different workers through strategies learned from data obtained during the interaction. This process reflects the principles of shared intentionality, as it involves a mutual understanding between the human and the robot about the worker's skills and preferences. It also illustrates the principle of interdependence, as the architecture relies on both the human and robot's capabilities to ensure safe and efficient

human-robot collaboration.

More concretely, DAC-HRC enables robotic components to tailor their interactions to the needs of their collaborative human partner, taking into account their unique skills, capabilities, and preferences. In order to achieve such a level of personalized adaptation to each human partner, each of its core four functionalities, control, safety, adaptation, and interaction are all distributed across the whole architecture, while having their specialized cognitive modules: the Task Planner, the Moral Engine, the Worker Model and the Interaction Manager, respectively.

DAC-HRC follows the design principles of the Distributed Adaptive Control theory, which states that the goal of cognition in embodied agents is to control action, and as such, any cognitive system can be described as a modular, hierarchical control system operating at different spatiotemporal timescales (P. F. Verschure, 2012). The DAC theory can be expressed as a robot-based cognitive architecture organized in two complementary structures: layers and columns. The columnar organization defines the processing of states of the world, the self, and the generation of action. The organizational layers define the different levels of control, starting from the Soma Layer integrating all sensors and effectors of the system, the real-time reactive sensorimotor control in the Reactive Layer, the adaptive associative learning and allostatic control in the Adaptive Layer, up to abstract and symbolic manipulation and context-based control in the Contextual Layer.

DAC-HRC is organized following DAC's layered structure, where each of its specialized HRC cognitive modules is located at different levels of the layered architecture based on their spatiotemporal timescales of control and their informational and sensory abstraction, as we can see in Figure 5.2. In other words, the cognitive modules are strategically distributed throughout the architecture based on the specific temporal and spatial requirements for control, as well as the degree to which they process and abstract sensory information and relevant data.

Its Soma Layer is defined by the hybrid combination of synthetic and biological sensors and actuators that comprise the aCell, that is, the cameras, the tablet, the cobot, and the human worker. In its reactive layer, DAC-HRC incorporates the Interaction Manager and the Task Planner modules. The Interaction Manager is devoted to the human-machine interaction protocols necessary to orchestrate communication between the human and the cobot. The Task Planner is in charge of the adequate task allocation among the members of the collaborative entity. It sequentially organizes the disassembly tasks and controls

the correct turn-taking behavior between the human and the cobot. The Moral Engine, which is in charge of providing an additional layer of safety that adapts the security distances and robot speed to the particular preferences of the human partner and the current task context, spans both the reactive and adaptive layers. The same applies to the API component of the Worker Model, which deals with the real-time information related to the worker, as well as with the update of the Database. In the contextual layer, the Worker Model Database provides the system with an internal model of the human workers, storing in its long-term memory the past interactions between each user and the system, as well as relevant information for adapting the overall collaboration to the preferences of the human partner. The Interaction Learner, spanning both the contextual and adaptive layers, uses the contextual information to learn from past interactions with the system to adapt the options displayed by the Interaction Manager through the tablet device. In the following sections, we describe in detail the technical implementation of the cognitive modules of the DAC-HRC architecture.

5.2.0.1 Task Planner

The DAC-HRC's Task Planner module is conceived as a human-in-the-loop hierarchical finite state machine that encompasses all disassembly steps of all devices, as well as the error-handling protocols. The Task Planner (TP) has been developed to ensure robust orchestration of various components contributing to the disassembly of WEEE devices within the aCell system. The objectives of the TP are to coordinate the proper disassembly steps for each device, organize the disassembly procedure and robot-worker interleaving, centralize task-related information among the DAC-HRC modules, and implement safe and robust error-handling protocols. The TP reflects the principles of shared intentionality and interdependence, as it involves a mutual understanding between the human and robot about the sequence of tasks and reliance on each other's capabilities to complete these tasks. The TP integrates and coordinates low-level sensorimotor information (coming from the computer vision and robotic components of the aCell) with high-level information about the task and the interaction (coming from the upper control layers of the architecture). Therefore, within the TP's HFSM, we find states with different levels of abstraction and description. The Task Planner operates at 5 levels:

- Task Planner. This level corresponds to the highest level of abstraction, which contains the state machines (SM) of all 4 Devices.

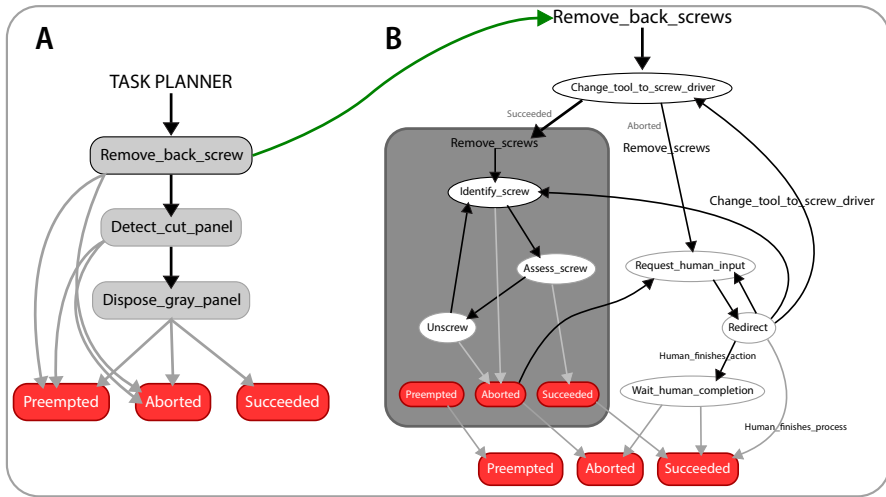


Figure 5.3. Task Planner. Visualization of the Device (a) and Task (b) levels of the Task Planner's hierarchical finite-state machine (HFSM) for the flat-panel display use case. **a)** Device level of the Task Planner's HFSM. This level showcases the finite-state machines responsible for sequentially connecting the disassembly steps (i.e., Tasks) required for the correct disassembly of the flat-panel display. **b)** Task level of the Task Planner's HFSM. At this level, dedicated error-handling mechanisms within the Task Planner engage the worker when errors occur during the execution of an action (in this case, 'remove screws'). This ensures that the worker is actively involved in resolving any issues that may arise. Feedback and responses from the worker serve to redirect the state of the Task Planner, enabling effective error resolution and maintaining the overall flow of the task execution.

It also contains the functionalities that deal with continuous status reports, as well as direct human interactions (through the Interaction Manager, or IM) that allow the TP to be suddenly interrupted by the worker.

- **Device.** This level contains the state machine (SM) that links the steps (i.e., Tasks) needed for the proper disassembly of a particular device, in a sequential manner (i.e., without internal loops). Thus, it comprises a straightforward sequential SM with all the necessary steps or Tasks to be executed in the right order, steps which have been pre-defined based on domain-specific knowledge of the proper disassembly of the devices (see Figure 5.3-A).
- **Task.** In this level, a particular Task –involving one or more Actions (see below)— is executed, with the end result of removing a particular component of the device (e.g., "top lid removal of the e-light"). Here, errors during the execution of an Action are handled in a dedicated SM so that the worker is engaged when-

ever needed (see Figure 5.3-B). Feedback and responses from the worker redirect the state of the TP accordingly (e.g., if an error with the robot occurs and the worker decides to complete that Task themselves).

- **Action.** This is the atomic level of description, where specific modules are uniquely engaged via ROS communication (e.g., ROS-actions or ROS services). During an Action, either a ROS action is sent to a robot to perform a specific action (e.g., “change tool to vacuum gripper”, or “dispose lid”), or a ROS service is issued to the vision module to acquire the necessary information that the robot will need to perform a subsequent action (e.g., “identify the grasping pose of the lid”).
- **Sub-action.** In some cases where Actions need to be repeated several times and imply feedback loops with vision and the robot, an additional level is introduced so that the SM design becomes more modular and robust (e.g., “Unscrew the 6 screws of the microwave’s cover” is designed so that a dedicated SM to unscrew coordinating the robot and vision feedback can be called in loop until all screws have been removed).

The Task Planner is implemented with the Smach-ROS python library, which allows seamless integration of HFSMs with ROS-based communication protocols (Bohren & Cousins, 2010; Pradalier, 2017). Crucially, internal data structures allow the conveying of information received from vision (response of a ROS service) to the robot (goal of a ROS action). In an SM, the transitions between states depend on the outcome of each State after having been executed. An example of a Task can be seen in Figure 5.3. In general, an outcome “succeeded” will make the SM transition to the next Task or Action (depending on the level). The outcome “aborted” will engage the error-handling loop (see section Error Handling below), which asks the worker for feedback, and transitions to different states according to the worker’s decision (e.g., the robot tries again, or the worker finishes the Task and then the TP moves to the next Task). The hierarchical structure of the FSM can be achieved because all SMs are treated as States too, inheriting their properties.

5.2.0.2 Interaction Manager and Interaction Learner

The Interaction Manager module plays a vital role in facilitating efficient communication and interaction between humans and robots. To achieve this, the module integrates multimodal channels of commu-

nication, ranging from audiovisual interfaces to embodied non-verbal communication. To account for high levels of noise and equipment worn by workers, verbal communication was excluded from the communication repertoire. The two main modes of interaction, gesture-based communication, and tablet-based communication, have been chosen to address the noisy industrial environment and safety concerns during collaboration between human workers and robots.

Gesture-based communication provides a fast and direct means for the human worker to convey simple and fast control commands and responses to the robotic companion, making it a useful and naturalistic way of communicating in the collaborative environment of the aCell. The Interaction Manager integrates eight different communication signals, providing a rich set of gestures for effective communication between the worker and the robotic system.

In contrast, tablet-based communication allows for more detailed and rich information exchange using images and text, enabling robot-to-human updates on the task's current state, system requests for human intervention, and human input for problem-solving or decision-making when required.

These two main modes of interaction have been chosen to cover the speed-accuracy trade-off, with gestures for simpler but time-sensitive interactions, and tablets for slower but more fine-grained information exchange. This dual communication paradigm accommodates individual human preferences and ensures efficient collaboration in various human-robot collaboration scenarios, as we will see in the Results section.

The Interaction Learner adds a level of personalization on top of the Interaction Manager functionalities by providing it with an adaptive mechanism to support human-robot decision-making based on the prior history of interactions between the human and the cobot. Its main function is to keep track of human-robot interactions and human feedback during error-handling scenarios. It computes useful statistics based on the history of human-robot interactions, and when a similar situation is encountered, it adapts the options displayed in the tablet by the Interaction Manager in a way that enhances collaborative decision-making by highlighting on the menu the most frequently selected options by that worker in a given situation. This level of adaptation takes into account the human-robot interactions at each specific step during the disassembly and for each worker in particular.

5.2.0.3 Moral Engine

The design of the Moral Engine incorporates a set of reactive control systems inspired by Pavlovian appetitive and aversive drives (Freire et al., 2020c). This approach shapes the Moral Engine's functionality, guiding its interactions in the DAC-HRC architecture to align with principles of both harm avoidance and proactive cooperation. This incorporation allows the Moral Engine to adapt its behavior dynamically, not only avoiding harm but also optimizing operation parameters such as speed, distance, and task allocation based on the unique context of each human worker. The Worker Model, integral to the contextual layer of the DAC-HRC architecture, helps personalize the interaction, treating each worker as a distinct individual with specific preferences and needs.

The Moral Engine module, in charge of providing a context-aware and personalized safety control system, spans across three layers of the DAC-HRC architecture. The Moral Engine's reactive layer integrates several homeostatic modules whose purpose is to monitor key aspects of the human-robot interaction. The goal of each homeostatic module is twofold: to keep its desired variable within the optimal range of operation, and to exert control when that variable trespasses the safe range. The current implementation comprises the homeostatic control of key proxemics variables in HRI, such as the human-robot interaction security distance, the robot movement speed, and the robot action execution speed. When any of those variables reach or trespass their threshold, the control response can be either a direct modification of the exceeded value -in the case of speed modulation-, or a command directed to stop the robot's current action -in case the HRI distances are trespassed. For instance, if the actual detected distance between the human and the robot is below the desired safety value, the homeostatic control system will generate a stop signal and the robot will not move until the actual distance goes back to the desired range.

The Moral Engine's adaptive layer is composed of the allostatic control module. This is the key mechanism by which the Moral Engine can adapt the interaction of the robot to its changing environment. This module is in charge of the transformation of the environmental information provided by the contextual layer and modifying the desired parameters of the subsequent homeostatic regulatory mechanisms of the reactive layer. For instance, when the robot is handling a dangerous tool, the allostatic control module gets this information and adapts the desired safety HRI distance, as well as the speed at which the robot

will operate when being close to a human.

The Moral Engine's contextual layer deals with the integration of the relevant environmental, social, and material information that comes from other modules of the DAC-HRC architecture. It endows the Moral Engine with context awareness. The constant integration of these different sources of information defines the specific context at every point in time, thus allowing the Moral Engine to monitor and adapt the behavior of the robot to the changing conditions of its surroundings. For instance, the contextual layer can obtain information in real time about the HRI preferences of the currently detected human worker, the risk level of the current robot action, and the information about the current tool being used by the robot (if any).

The incorporation of the reactive control systems inspired by the Pavlovian appetitive and aversive drives allows the Moral Engine to adapt its behavior dynamically, not only avoiding harm but also optimizing operation parameters such as speed, distance, and task allocation based on the unique context of each human worker. The worker model, integral to the contextual layer of the DAC-HRC architecture, helps personalize the interaction, treating each worker as a distinct individual with specific preferences and needs.

5.2.0.4 Worker Model

The Worker Model is composed of short-term and long-term memory buffers along with its reactive and adaptive input processing layers. The Worker Model's reactive layer serves as a first data integration step, gathering information from several input sources, whereas its adaptive layer processes the raw data in order to produce new parameters that will be used by other modules of the Worker Model and the DAC-HRC architecture. The online information gathered by the Worker Model's reactive layer is transiently stored in the short-term memory buffer before it is further processed by the adaptive Layer to generate new relevant information about the worker and their interaction with the system. For instance, the short-term memory can store the timings of past interactions during a disassembly step, while the adaptive layer generates estimates of current task duration based on this input. The type of input information gathered by the Worker Model can be divided into offline and online variables:

- Offline variables - This type of data is mostly static, as it will not vary throughout the session (e.g., age, gender, language, and interaction preferences). This information is acquired through preliminary questionnaires before engaging with the system and defines the profile of each user based on demographic information and their opinion towards robots.
- Online variables - Comprises all the relevant user data that is dynamically updated in real-time over the course of the interaction with the system. Integrates information about the position of the worker and their performance (e.g., current disassembly task, or estimated duration), as well as about the context in which the worker is embedded (e.g., current disassembly process, aCell number, or location).

The technical implementation of the Worker Model is based on two main components: the Worker Model's API and the Worker Model Database. The database implements the long-term memory component of the Worker Model. Its function is to store all the information related to each worker and to keep it up to date. It is deployed as a document-oriented database using MongoDB ¹, where each worker profile is stored as a unique document. Each worker model entry is initialized with the offline variables acquired from the worker profile and questionnaires. Additionally, it also stores the main statistics of each interaction between the worker and the DAC-HRC collaborative architecture that has been extracted by the Worker Model API, such as the expected task duration or the history of interactions with the tablet.

All the communications with the database are centrally controlled by the Worker Model's API, which integrates the reactive and adaptive input processing layers along with the short-term memory component of the Worker Model. The API's function is twofold: it performs the basic CRUD (create, read, update, and delete) operations that keep the database up to date, and it is in charge of filtering the online and state variables to produce the task- and interaction-relevant outputs of the Worker Model. The API is written in Python and communicates with the database using BSON as the data interchange format.

¹<https://www.mongodb.com>

5.3 Results

Turn-taking Human-Robot collaborative interaction in the disassembly of a WEEE device

This use case describes the involvement of the DAC-HRC architecture during the collaborative disassembly of WEEE devices between a cobot and a human worker. Such a collaborative process begins when a human worker approaches the aCell. Once the worker enters the working area, they are recognized by the vision module that perceives their identity by decoding the unique fiducial code allocated in the workers' helmets (Fig. 5.4a). Then, using the identity of the worker, the Worker Model anonymously accesses their corresponding personal information and makes it available to the entire DAC-HRC architecture, so other cognitive modules can socially adapt to the current worker. This process reflects the principle of shared intentionality, as it involves a mutual understanding between the human and the cobot about the identity and role of the current worker. It also illustrates the principle of interdependence, as the overall disassembly performance depends on both the cobot and worker (Fig. 5.4b).

The Interaction Manager receives and processes the worker's personal information and provides them with immediate feedback about their detection by displaying such information through the IM app (Fig. 5.4c). It is noteworthy that this information, and further notifications, are displayed meeting the worker language's preferences. Importantly, the rapid communication between the vision modules and the Worker Model ensures that the social information considered by the cognitive architecture has real-time correspondence with the current human worker at the aCell.

In parallel, the cobot continues operating primarily guided by the goals imposed by the Task Planner. The succession of steps, as well as their status and the progress during the disassembling process, is also communicated to the human worker through the IM app (Fig. 5.4-c). However, as mentioned in the description of the Task Planner, the scheduling of disassembly steps is determined as a succession of states that ensures the task allocation (human or cobot) matches the worker's skills and preferences. Thus, the optimal distribution of disassembly tasks leads to stable collaborative turn-taking dynamics, fostering predictability and facilitating the rapid acquisition of social conventions



Figure 5.4. Human-Robot Collaborative Disassembly of a WEEE Device. (a) Vision module identifies a worker using their unique code. (b) DAC-HRC architecture adjusts to worker preferences, modulating robot behavior. (c) The IM app shows worker details and disassembly status, and sends notifications if human input is required. (d) Task Planner updates after full disassembly of the WEEE device.

(Freire et al., 2020b; Hawkins & Goldstone, 2016).

Once the Task Planner has successfully overcome the robot's assignments and reached an action that requires human intervention, this module interplays with the Interaction Manager to proactively interact with the human worker. As a result, the IM app sends a notification to the human worker describing the action to be performed (Fig. 5.4-c). Moreover, this notification enables the worker to control the clamping tool (see Fig. 5.1) through the IM app when the device's translation or reorientation is needed. Once the human intervention has been completed, a completion button must be pressed to allow the DAC-HRC architecture to carry on with the next step. Additionally, an abort option is available in cases where the human worker needs to stop the collaborative disassembly and finish on their own.

Finally, when both the robot and human's disassembling actions have been completed, the Interaction Manager, in liaison with the Task Planner, informs the human worker about the completion of the disassembling process through the IM app (Fig. 5.4-d).

Error handling mechanisms personalized by past collaborative interactions

Given the challenge that the disassembling of WEEE devices represents and the complex interaction that DAC-HRC cognitive architecture maintains within its components and with the external HRR sensory and control modules, there is the possibility of errors arising during the disassembly. These errors may prevent the optimal performance of the cobot and thus the successful execution of the disassembly steps (Fig. 5.5-a).

To overcome these errors, the Task Planner, through the IM app, informs the human worker of any problematic action (i.e., any action that leads to errors) and provides three possible solutions. These options give the worker the possibility to (1) force the cobot to retry the problematic action, (2) inform the cobot that the worker will take care of the action (Fig. 5.5-b), or (3) to inform the cobot that the worker will take care of the remaining steps of the disassembly (Fig. 5.5-c).

Importantly, due to the involvement of the Interaction Learner module, this error-handling functionality becomes adaptive to the worker's preferences by learning from previous interactions. Thus, if a worker



Figure 5.5. Personalized error-handling mechanisms during Human-Robot collaborative disassembly. **a)** The complex coupling of both the aCel and the DAC-HRC architecture becomes a potential source of failure that needs to be addressed at the systems level. **b)** Either when the cobot cannot complete a given disassembly action, or when is the worker's turn to execute a step of the disassembly, the human worker can intervene safely. **c)** IM app notification of an error during the disassembly providing the three different actions to overcome the error. **d)** The IM app in liaison with the Interaction Learner provides an attentional bias towards the preferred error-handling options by modulating their visual saliency.

exhibits consistent biases toward one of the options when handling Task Planner errors, the module memorizes these preferences and facilitates future decision-making by increasing the visual saliency of the previously preferred options (Fig. 5.5-d). This reflects the principles of shared intentionality and interdependence, as it involves a mutual understanding between the human and robot about handling errors and reliance on each other's capabilities to resolve these errors and complete the disassembly process.

Adaptive and personalized safety measures for Human-Robot collaboration

In parallel to the Human-Robot collaborative disassembling of WEEE devices, safety-related information is constantly monitored and processed to provide adaptive and personalized robot behavior. With this aim, once the computer vision module has detected and recognized a human worker at the aCell, the Moral Engine draws its measure of trust from the Worker Model. In addition, the Moral Engine updates the safety distance and robot's speed according to the worker's preferences (Fig. 5.6). This process reflects the principles of shared intentionality and interdependence, as it involves a mutual understanding between the human and robot about the worker's trust level and reliance on each other's capabilities to maintain safety during the disassembly process.

Since the adaptation of the robot's speed to the worker's trust does not interfere with the worker's performance, it has been designed to occur covertly and automatically. Thus, the robot's speed is set at high levels when the current worker has reported high levels of trust in their robotic counterpart and decreases when a more distrusting worker enters the aCell.

However, aiming to ensure the integrity of the human workers, the normal turn-taking collaborative Human-Robot interaction can be interrupted when they surpass the safety distance (Fig. 5.6-a-b). This safety distance, as well as the robot's speed, is initially personalized by the Moral Engine based on the trust information provided by the Worker Model. Thus, human workers with higher trust are allowed to get closer to the workbench while the cobot is carrying out its tasks. Nonetheless, when a more distrusting worker enters the aCell, this safety distance is enlarged ensuring both their physical integrity and physiological well-being (Fig. 5.6-c-d and Supplementary Video 1). Importantly, even when workers report a maximum level of trust in robots, a minimum safety distance is set ensuring European safety standards (ISO, 2011, 2016). The real-time monitoring of the workers' position relative to their personalized safety distance is accomplished by the DAC-HRC architecture due to the continuous communication between the Moral Engine and the vision module, which provides the current worker's location.

In cases where the worker has surpassed their personalized safety distance, the Moral Engine ensures their integrity by immediately stop-

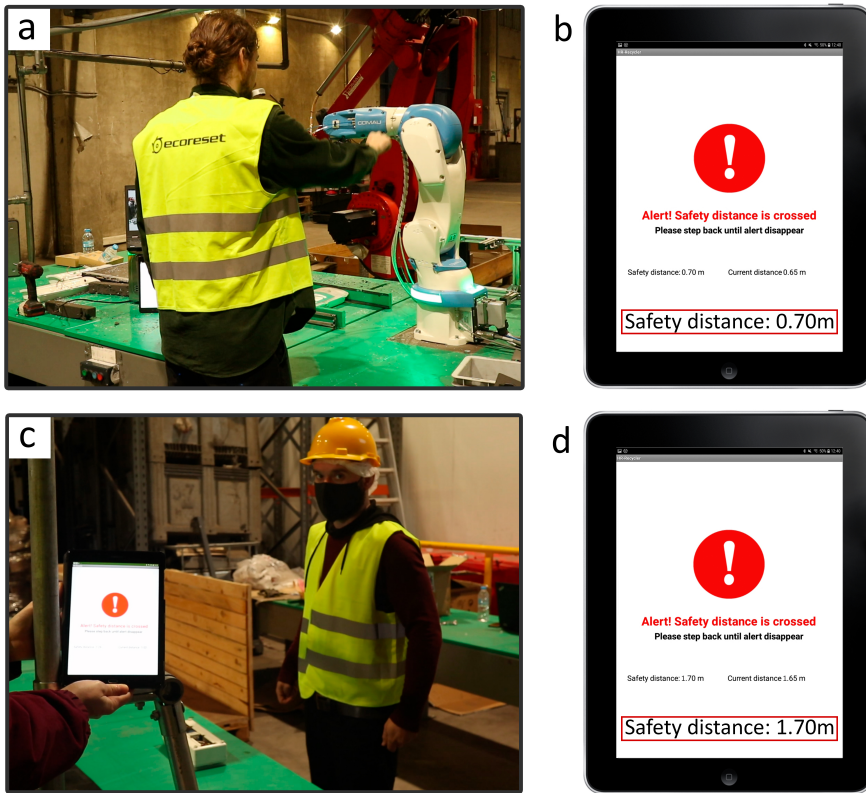


Figure 5.6. Adaptation of safety distance to human workers with different trust measures. **a)** A worker with high trust in their robotic counterpart engages in the collaborative process of disassembling the WEEE device. Their high trust is considered by the Moral Engine which, accordingly, sets a short safety distance. Nonetheless, once this personalized safety distance is surpassed the robot comes back to its initial position and stops. **b)** Surpass of the safety distance triggers an IM app alert notification. According to the worker's high measure of trust, the personalized safety distance is set at 0.70 meters. **c)** When a different worker reporting a lower measure of trust in their robotic counterpart enters the aCell, the Moral Engine recalculates the safety distance. As a result, the safety distance is enlarged and the human worker is not allowed to get so close to the cobot without triggering the safety alert. **d)** IM app alert notification when the worker with lower trust surpasses the safety distance. Notice that it was enlarged to 1.70 meters.

ping the cobot's action. The trespassing of the safety distance is also reported to the Interaction Manager, which in turn notifies the human worker about their current location and the minimum distance they should keep to the cobot (Fig. 5.6-b and c). This alert remains displayed on the IM app until the worker gets back to respect their safety

distance. Once the safety distance is reached again, the Moral Engine's alert disappears from the IM app and the cobot resumes its previous task.

Gesture-based communication for goal-oriented collaboration

Besides the direct input that human workers could provide to the DAC-HRR architecture through the IM app, vision modules recognize a set of gestures that enable multimodal communication and enhance human-robot interaction during collaborative disassembly.

Unlike direct input through the IM app, which is dependent on specific events such as the requirement of human intervention or error-handling situations, gesture-based communication is available at any time during disassembly. That is, the workers can exert control over the collaborative process by performing predefined gestures that inform the DAC-HRC architecture to stop or resume the disassembling process, as well as informing that the disposal tray is full (Fig. 5.7 and Supplementary Videos 2 and 3). Consequently, the worker also gets feedback about the detection of the recognized gesture through the IM app (Fig. 5.7-b, d, and e). This reflects the principles of shared intentionality and interdependence, as it involves a mutual understanding between the human and robot about the meaning of different gestures, and a reliance on each other's capabilities to interpret these gestures and respond appropriately.

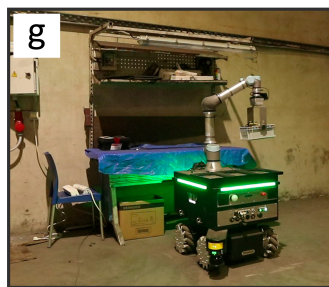
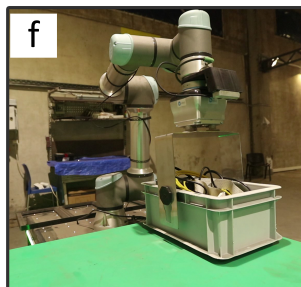
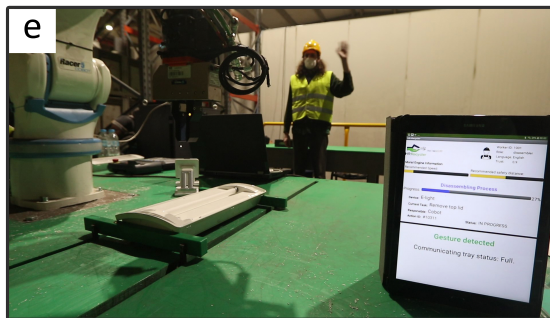
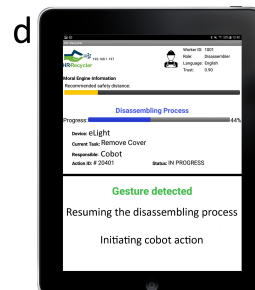
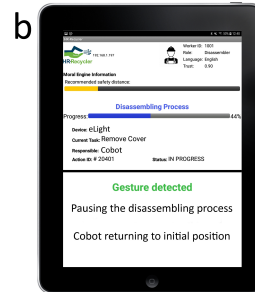
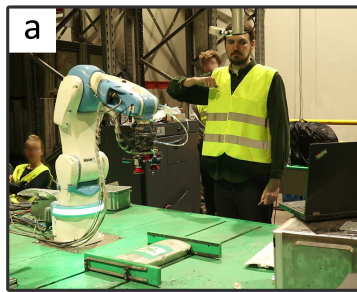


Figure 5.7. Human-Robot interaction based on gesture recognition. **a)** Human worker performing the 'stop' gesture. **b)** IM app notification for the recognition of the 'stop' gesture and showing information about the corresponding robot action: adopting its initial, default pose. **c)** Human worker performing the 'resuming' gesture. **d)** IM app notification for recognition of the 'resuming' gesture and showing information about the corresponding robot action, resuming the interrupted action. **e)** Human worker performing the 'wave' gesture and IM app notifying about the recognition and meaning of the gesture: disposal tray is full. **f)** Automated guided vehicle (AGV) robot picking up the full disposal tray from the aCell. **g)** AGV robot leaving the full disposal tray in the removal area.

5.4 Discussion

This chapter presents a novel cognitive architecture, DAC-HRC, based on the Distributed Adaptive Control theory, aimed at controlling the interaction of a collaborative entity, the aCell, composed of a human and an industrial cobot. DAC-HRC consists of several cognitive modules operating at different timescales and levels of abstraction, working together to produce adaptive collaboration that is personalized to each human user. The main cognitive modules of the architecture are the Task Planner, Moral Engine, Interaction Manager, and Worker Model.

With this work, we contribute to the development of novel human-robot collaboration paradigms tailored to the specific requirements of Industry 4.0, as previous human-robot interaction paradigms have remained more rigid and static in industry settings. More concretely, we apply novel HRC schemes to non-humanoid robots in industrial contexts, such as gesture-based communication schemes and adaptation to the user and context, providing more organic human-robot interaction. The DAC-HRC architecture is applied to a specific use case: the HR-Recycler environment (Axenopoulos et al., 2019), a hybrid recycling plant focused on the disassembly and recycling of WEEE devices.

We showcase the capabilities of the architecture in several pilot studies, highlighting the different DAC-HRC functionalities that enhance human-robot collaboration in the recycling plant, concretely: (1) turn-taking human-robot collaborative interaction during the disassembly of a WEEE device, (2) error handling mechanisms personalized by past collaborative interactions, (3) adaptive and personalized safety measures for human-robot collaboration, and (4) gesture-based communication for goal-oriented collaboration. Each of the modules of the DAC-HRC architecture covers different aspects of the collaboration be-

tween the human and robot, such as adaptation, safety, personalization, transparency, and real-time interaction, among others.

The modular design of the DAC-HRC architecture benefits human-robot collaboration by providing flexibility and adaptability to different contexts and requirements. Architectures like DAC-HRC can have a significant impact on Industry 4.0 and the field of human-robot interaction and collaboration, as they enable more seamless and efficient collaboration between humans and robots in various industrial settings.

Current limitations of the DAC-HRC architecture include the need for further validation and refinement across diverse industrial settings, as well as the integration of additional cognitive modules to address other aspects of human-robot collaboration. Examples of such modules may include those focusing on predictive and adaptive task allocation, human action understanding, emotion recognition, and fatigue detection.

Therefore, future work should focus on expanding the range of applications and use cases for the DAC-HRC architecture, such as in healthcare for patient assistance or in logistics for warehouse management. Additionally, the development of more advanced techniques for real-time adaptation and learning from human-robot interactions is needed, like machine learning algorithms that can better predict human behavior and intentions (Freire et al., 2023; Lake et al., 2017; Shum et al., 2019) or natural language understanding methods to enhance communication and coordination between human and robotic workers (Dong et al., 2019; Thomason et al., 2019).

In terms of shared intentionality and interdependence, the DAC-HRC architecture can inform future work on human-robot collaboration in industry and beyond by fostering a deeper understanding of the cognitive and communicative processes involved in collaborative tasks. This understanding can lead to the development of more advanced and adaptive hybrid collaborative systems that better support human-robot interactions in a variety of contexts. For instance, in a manufacturing setting, this understanding could lead to the development of cobots that can better anticipate the needs of human workers, adjusting their actions in real-time to assist when and where it's most needed. This could involve anything from passing tools to a worker at the right moment to adjusting the pace of a production line based on the worker's current workload and fatigue. In a healthcare context, this could result in assistive robots that are more responsive to the

needs of patients or healthcare providers based on their capacity to learn from previous interactions. For example, a robot could recognize when a patient is struggling with a task and offer assistance, or it could adapt its behavior based on the preferences and perceived emotions of different patients.

The collaborative entity composed of two agents (in our case, a human and a robot), the aCell, can be likened to a distributed cognitive organism, such as an ant colony. This comparison is rooted in the notions of extended cognition (Clark & Chalmers, 1998) and liquid brains (Solé et al., 2019), which describe how cognitive processes can be distributed across multiple agents in a system, rather than being confined to a single individual.

In an ant colony, for example, no single ant possesses the entire knowledge of the colony's activities. Instead, each ant contributes to the collective intelligence of the colony through its individual actions and interactions with other ants. Similarly, in an aCell, the human and robot work together as a cohesive unit, with each contributing their unique skills and capabilities to the collective performance of the task at hand.

This perspective offers valuable insights for designing distributed hybrid collaborative systems. For instance, it suggests that we should focus not only on the individual capabilities of humans and robots but also on how they can best interact and coordinate their actions to achieve shared goals. This could involve developing communication protocols that enable humans and robots to share information more effectively or designing control algorithms that allow robots to adapt their behavior based on the actions of their human partners. By capitalizing on the complementary strengths of humans and robots in this way, we can create hybrid collaborative systems that enable them to work together more effectively and efficiently.

In conclusion, the DAC-HRC cognitive architecture represents a significant step forward in the field of human-robot interaction and collaboration, particularly within the context of Industry 4.0. By integrating novel HRI schemes and addressing various aspects of collaboration, such as adaptation, safety, personalization, and transparency, the architecture provides a flexible and modular framework for developing more effective and efficient human-robot collaborative systems. The insights gained from this work have the potential to significantly impact not only the future of industrial human-robot collaboration but also the broader field of human-robot interaction research.

Acknowledgements

This study was funded by the Hybrid Human-Robot RECYcling plant for electriCal and eLEctRonic equipment Horizon 2020 project — HR-RECYCLER (ID: 820742). We would like to thank our colleagues in HR-Recycler since the present publication highly depends on the integration with other partners' contributions. Specially, we would like to thank the efforts and time spent by Albert Tissot, who not only lent us his image for our pictures and videos but also greatly helped us in integrating the HRC-DAC architecture with the vision modules and deploying our models in real scenarios.

6

CHAPTER 6

Unifying Hippocampal-VTA Roles during Goal-Directed Behavior

ABSTRACT

The mammalian hippocampus is central to navigation, memory, and novelty detection. In this chapter, we revisit and extend the hippocampal-VTA loop model to propose a unified framework for understanding hippocampal function during goal-directed behavior. Specifically, we describe how key hippocampal features, including (i) the emergence of place cells, (ii) the stabilization of cognitive maps, (iii) the clustering of place fields near reward sites, and (iv) the modulation of hippocampal replay, can be more comprehensively understood when analyzed through their interplay with the ventral tegmental area (VTA). To explore this framework, we outline a biologically constrained cognitive architecture capable of reproducing key hippocampal-VTA phenomena, including the emergence and reorganization of place cells from egocentric inputs, novelty-driven hippocampal activation, and the enhancement of sharp-wave ripples (SWRs) in response to novelty. Preliminary computational experiments provide encouraging evidence that the hippocampal-VTA loop operates as a dynamic mechanism for integrating spatial, motivational, and novelty-related information, enabling adaptive behavior in complex and changing environments. While still at an exploratory stage, this framework sets the foundation for future experimental and theoretical work aimed at unifying hippocampal and VTA contributions to adaptive autonomy.

This chapter builds upon work presented in the previous chapters to outline future research directions. It provides an extensive review of relevant literature, formulates specific hypotheses, and presents preliminary findings. However, this line of research is still in progress and has not yet reached the level of maturity required for definitive scientific conclusions or publication in peer-reviewed venues. Accordingly, the chapter should be read as a proposal for a research agenda aimed at guiding future investigations into the hippocampal-VTA loop and its role in goal-directed behavior.

6.1 Introduction

Classical clinical cases in neuropsychology first revealed the hippocampus as central to memory formation. A seminal example is the case of Henry Molaison (H.M.), whose severe epilepsy led to a bilateral medial temporal lobectomy in 1953, removing two-thirds of his hippocampus (Scoville & Milner, 1957). This radical intervention resulted in profound amnesia: H.M. was unable to recall experiences from the two years preceding the surgery and could no longer form new long-term memories. Remarkably, the deficits extended beyond declarative memory into the spatial domain. H.M. was unable to remember what he had eaten only minutes earlier, nor could he navigate to his family's new home after moving. These clinical observations underscored the hippocampus as a structure crucial not only for episodic memory but also for navigation.

Guided by Tolman's theory, which proposes that the mammalian brain processes sensory information to construct internal cognitive maps of the environment (Tolman, 1948), and on experimental reports linking hippocampal lesions to impaired performance in spatial tasks (Douglas, 1967), John O'Keefe made groundbreaking discoveries about hippocampal coding. He identified **place cells** in the CA1 layer of the hippocampus that fired selectively when an animal occupied a specific location in its environment (O'Keefe, 1976; O'Keefe & Dostrovsky, 1971). Each place cell encoded a distinct spatial region, collectively forming a map-like representation of space. O'Keefe further described **misplace cells**, which responded when expected and actual sensory inputs diverged (O'Keefe & Dostrovsky, 1971). This finding illuminated a third critical function of the hippocampus: the detection of novelty through mismatches between prediction and experience.

Building on these discoveries, John Lisman & Antony Grace proposed that the hippocampus and ventral tegmental area (VTA) form a reciprocal loop in which novelty signals originating in the hippocampus are conveyed to the VTA, which in turn facilitates the familiarization of novel events through dopaminergic feedback to the hippocampus (Lisman & Grace, 2005). This hippocampal-VTA loop can be summarized in five interconnected steps:

1. **Prediction and Comparison:** Sensory input from layer II of the entorhinal cortex (EC) reaches the hippocampal dentate gyrus and CA3 region, which generate memory sequences predicting

potential future events. These predictions are transmitted via Schaffer collaterals from CA3 to CA1, where they are compared with current sensory input provided directly from EC layer III (Lisman, 1999).

2. **Novelty Signal Generation:** If a mismatch is detected, a novelty signal is generated as a fast (<100 ms) and robust activation of CA1 cells (Ruusuvirta et al., 1995). This signal is relayed through the subiculum (Sub) to the nucleus accumbens (NAcc) (Floresco et al., 2001; Ihalaainen et al., 1999; Legault et al., 2000) and ventral pallidum (VP) (Chrobak & Napier, 1993; Mogenson et al., 2019), ultimately releasing VTA dopaminergic neurons from tonic inhibition (Floresco et al., 2003).
3. **Dopaminergic Feedback:** As result of disinhibition, VTA releases dopamine in CA1 facilitating long-term potentiation (LTP) (Frey, Schroeder, et al., 1990; Frey et al., 1991; Huang & Kandel, 1995) and closing the hippocampal-VTA loop (Gasbarri et al., 1997).
4. **Integration of Goal-Dependent Signals:** The prefrontal cortex (PFC) provides goal-related input that further informs the loop (French & Totterdell, 2002). Accumbens spiny cells, which exhibit bistable membrane potentials (O'Donnell & Grace, 1995), are particularly sensitive to PFC excitation when their UP state has been primed by hippocampal NMDAR activation (O'Donnell & Grace, 1998).
5. **Affective, motivational, and saliency information:** The pedunculo-pontine nucleus (PPT), via glutamatergic projections, modulates VTA dopaminergic activity, providing affective, motivational, and saliency-related influences (Kobayashi et al., 2002; Koyama et al., 1994; Semba & Fibiger, 1992). Notably, simultaneous activation of the subiculum and PPT markedly increases the number of active dopaminergic neurons, amplifying the signal threefold (Lodge & Grace, 2006).

Two decades after its proposal, the hippocampal-VTA loop remains the prevailing model for understanding novelty detection. Subsequent evidence has further substantiated its core mechanisms. For instance, the hippocampus not only exhibits a generalized increase in CA1 activity during exposure to novel contexts (Larkin et al., 2014). It also exhibits a phase shift in cell firing preference to later stages of the theta cycle (Lever et al., 2010). Moreover, when task contingencies change, modulation of VTA dopaminergic neurons by unexpected rewards has been shown to depend on intact hippocampal function (Takahashi et

al., 2024). Yet, despite these advances, the original loop does not fully integrate the growing body of findings on hippocampal dynamics, reward-related modulation, and replay mechanisms. These emerging insights call for a revised framework that extends the classical model beyond novelty detection toward a more comprehensive account of hippocampal contributions to goal-directed behavior.

6.2 A proposal for updating the HPC-VTA loop model

In light of more recent findings, we propose a revised framework for the hippocampal-VTA loop that extends beyond novelty detection to incorporate learning and memory processes essential for goal-directed behavior. This framework seeks to account for phenomena that the original model left unexplained, including the stabilization of place cell firing fields, the occurrence and role of sharp-wave ripples (SWRs), and the clustering of place fields near reward sites. In the following, we outline this extended model in five interconnected steps:

1. **Diffuse Cognitive Maps Signal Novelty:** The first step in the revised hippocampal-VTA loop emphasizes CA1's role in detecting novelty by comparing predictions with incoming sensory input. We propose that the novelty-driven increase in CA1 activity (Rusuvirta et al., 1995) reflects the lack of a reliable cognitive map. In a novel environment, or in a familiar one where contingencies have changed, CA3 cannot generate accurate predictions. As a result, CA1 receives mismatched inputs, leading its place cells to exhibit broad, overlapping firing fields (Fyhn et al., 2002; Kim et al., 2020). This diffuse activation produces widespread CA1 activity, which serves as a novelty signal to downstream structures and primes the hippocampal-VTA loop for subsequent memory formation and learning.
2. **Dopaminergic Modulation of Sharp-Wave Ripples (SWRs):** Following the pathway described by Lisman & Grace (Subiculum → Nucleus Accumbens → Ventral Pallidum), CA1 novelty signals release VTA dopaminergic neurons from tonic inhibition. The VTA then sends dopamine back to CA1, where it not only facilitates long-term potentiation (LTP) but also enhances hippocampal sharp-wave ripples (SWRs) by promoting their frequency and intensity (Gomperts et al., 2015; Kleinman & Foster, 2024; McNa-

mara et al., 2014; Miyawaki et al., 2014). During SWRs, place cells involved in recent experience reactivate, preserving their sequential order (i.e., replay). A process found critical for refining spatial representations, consolidating memory, and stabilizing firing fields across repeated exposure (Roux et al., 2017). This stabilization mechanism explains why novelty-driven CA1 activation fades with familiarity: as SWR-mediated plasticity consolidates experience, the hippocampus transitions from an exploratory to a stable representational state (Fyhn et al., 2002).

3. **Integration of Novelty and Reward Signals:** When a novel object is encountered and judged motivationally relevant, two disinhibitory pathways converge on the VTA, enhancing dopamine release. As previously described, the first arises from CA1 novelty signals. The second pathway originates in the hypothalamus: if the object fulfills a physiological need, hypothalamic input further suppresses GABAergic interneurons in the VTA, amplifying the dopaminergic response (Grove et al., 2022). This second pathway not only explains how VTA dopaminergic signals encode reward magnitude (Schultz et al., 1997; Tobler et al., 2005) but also how hippocampal replays do it as well (Ambrose et al., 2016). Together, this dual mechanism strengthens memory formation and integrates spatial and motivational information to support adaptive behavior (Kennedy & Shapiro, 2009).
4. **Reward Overrepresentation and Dopamine Ramping:** Novelty-driven dopaminergic signals, when amplified by motivational relevance, trigger a remapping process in CA1 in which place fields cluster around reward sites. This phenomenon is known as reward overrepresentation (Dupret et al., 2010; Hollup et al., 2001). This spatial clustering may function as a predictive signal: as the animal approaches a reward, an increasing number of place cells become active, further releasing VTA dopaminergic neurons from inhibition. Such dynamics could underlie the ramping activity observed in VTA dopamine cells during reward approach (Farrell et al., 2022). Interestingly, ramping signals have also been reported in the orbitofrontal cortex (OFC) and ventral striatum (vStr) (Basu et al., 2021; Van Der Meer & Redish, 2009). These three areas not only receive hippocampal input (Barbas & Blatt, 1995; Finch, 1996; Lisman & Grace, 2005; Wikenheiser & Schoenbaum, 2016) but synchronize their ramp cell activity with hippocampal theta oscillations (Fujisawa & Buzsáki, 2011; Van Der Meer & Redish, 2011; Young & Shapiro, 2011).
5. **Cognitive Map Update following Reward Omission:** The

hippocampal-VTA loop gains flexibility through its interaction with the lateral habenula (LHb), which encodes reward prediction errors in opposition to VTA dopaminergic neurons (Hu et al., 2020; Matsumoto & Hikosaka, 2007). LHb activity increases when an expected reward is omitted and decreases when an unexpected reward is delivered. Through GABAergic neurons in the rostromedial tegmental nucleus (RMTg), this pathway strongly suppresses VTA dopaminergic firing (Balcita-Pedicino et al., 2011; Jhou et al., 2009), a response tuned specifically to reward omission rather than to aversive events (Tian & Uchida, 2015). Reduced dopaminergic feedback to CA1 may prevent stabilization of the current cognitive map (Krishnan et al., 2022; Mizumori & Baker, 2017), instead biasing the hippocampus toward exploratory activity or map remapping when contingencies change. This mechanism ensures that the hippocampal-VTA loop remains responsive to environmental volatility and supports behavioral flexibility. Consistent with this view, inactivation of the LHb disrupts the ability to update or abandon ineffective strategies, underscoring its critical role in adaptive learning (Stopper & Floresco, 2014; Thornton & Evans, 1984).

6.3 Computational validation of the HPC-VTA model

Testing the hypotheses derived from theoretical frameworks directly in physiological or behavioral studies is technically challenging and often impractical. In this section, we offer preliminary computational validation to assess the framework's plausibility. We first examine whether a hippocampal autoencoder reproduces key signatures of novelty and stabilization (diffuse-to-focused CA1-like fields, reconstruction dynamics, and map similarity across trials and environments). We then ask whether reward-biased replay yields reward-overrepresented cognitive maps and whether such maps improve behavior by integrating them into a reinforcement-learning agent. Taken together, these simulations provide an initial test bed for the framework and generate concrete predictions for future physiological studies.

Hippocampal novelty signals can be derived from the diffuse firing fields of CA1 cells.

Our first hypothesis posits that novelty signals in CA1 emerge from the hippocampus's inability to form a stable cognitive map when environmental predictions are unreliable. In a novel environment, CA3 fails to generate accurate expectations, while CA1 continues to receive current sensory input from the entorhinal cortex (EC) layer III. The resulting mismatch between predicted and actual sensory information prevents CA1 from establishing well-defined place fields, resulting in broad, overlapping activity patterns that signal novelty.

To examine this hypothesis, we trained a hippocampal autoencoder across three visually distinct environments, each configured as a T-maze of identical size and geometry (see Fig. 4.4). The model was trained sequentially: first in T-maze v1, then in T-maze v2, and finally in T-maze v3, with 100 training trials per environment in a continuous learning regime. Several quantitative measures collected during training indicate that the model captures the proposed novelty detection dynamics of CA1 activity.

We first analyzed the mean spatial coverage of the activation patterns produced by the latent units in our hippocampal autoencoder. Although the model is non-spiking, each latent unit's activation can be visualized as a firing field-like pattern, reflecting the regions of the environment where that unit is most active. During training in the initial environment (T-maze v1), these activation fields were large and diffuse, often spanning most of the arena (mean coverage $\approx 85\%$; Fig. 6.1a). As learning progressed, the mean area covered by each unit's field decreased sharply and eventually stabilized, indicating the emergence of spatially selective representations. When the model was subsequently exposed to a novel environment (T-maze v2), mean coverage increased abruptly during the first trials, reflecting a transient loss of spatial selectivity. With continued training, coverage gradually decreased again but remained considerably higher than in the first environment. A similar pattern was observed when the model transitioned to T-maze v3.

This modulation of firing field coverage was accompanied by transient peaks of overall activity when examining the summed activation of all latent units at representative spatial locations (maze alleys and their intersection). Analogous to the transient increase in hippocampal activity that drives VTA disinhibition during novelty exposure, our model reproduced elevated activation levels at specific locations during the

early training trials of each new environment (Fig. 6.1b).

An individual analysis of a few example units can provide a more detailed idea of how the hippocampal autoencoder signals novelty. In Fig. 6.1c, we can observe how 5 example units initiate the training (T 0) at T-maze v1 displaying large firing fields. At trial 10 (T 10), these firing fields have been drastically constrained to specific locations resembling place cell-like activity. The firing field constraint continues until the end of the training session in T-maze v1. Then, when the model is presented with a novel environment (T-maze v2), these example units show two different behaviors. First, they remap, showing a spatial preference distinct from that developed in T-maze v1. Second, firing fields are initially larger. With training, this spatial preference stabilizes and constrains until the model is presented with a new environment.

A closer inspection of individual latent units provides a more detailed view of how the hippocampal autoencoder signals novelty. As shown in Fig. 6.1c, five representative units display broad activation fields at the beginning of training (T0) in T-maze v1, covering large portions of the environment. By trial 10 (T10), these activation fields become spatially constrained, resembling place cell-like patterns. This spatial selectivity is maintained until the end of the training session in T-maze v1. When the model is subsequently exposed to a novel environment (T-maze v2), the same units exhibit two characteristic responses: (1) Remapping, expressed as a shift in spatial preference relative to the fields formed in T-maze v1; and (2) field broadening, where activation patterns temporarily expand before stabilizing again with continued training. These dynamics illustrate how the model captures novelty-induced reorganization of hippocampal representations, consistent with empirical observations of CA1 remapping during environmental change. A similar novelty responses were observed when the model transitioned to T-maze v3.

These dynamics in latent unit activity were closely synchronized with the model's reconstruction performance (Fig. 6.2a). As training progressed, the autoencoder gradually minimized the input-output difference (reconstruction loss). However, when a model trained in one environment was exposed to a novel one, reconstruction loss increased sharply, reflecting the temporary inability to accurately encode the novel experience.

A comparable pattern was observed in the stabilization of the model's cognitive map (Fig. 6.2b). To quantify this effect, we computed Pearson correlations between each unit's activation map (firing field-like

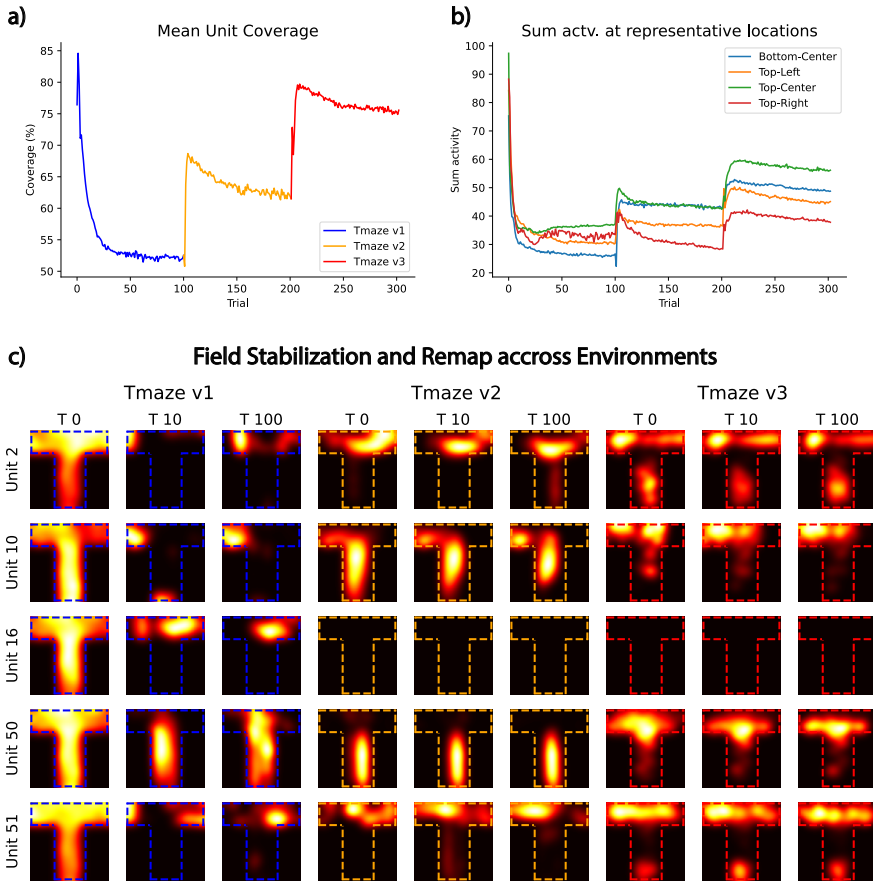


Figure 6.1. Hippocampal autoencoder signals novelty through diffuse activation fields. **a)** During initial training, latent vector units exhibit broad activation fields covering a large portion of the environment ($\approx 85\%$). With continued exposure to the same environment, these fields become spatially constrained, indicating increased selectivity. When the model is introduced to a novel environment, average field coverage rises sharply, reflecting the breakdown of previously stabilized representations. **b)** Overall latent vector activity measured at representative spatial locations (maze intersection and alley ends) follows a similar dynamic to unit field coverage, with transient peaks during the early trials of novel environments. **c)** Example activation fields of five latent units across training trials and environments. Each unit develops localized, place cell-like selectivity that remaps and broadens upon exposure to a new environment before stabilizing again with experience.

pattern) at trial t and $t + 1$. Consistent with the previous analyses, these correlations showed rapid stabilization within the same environment and pronounced remapping upon transition to a new one.

Both effects, the reduction in reconstruction loss and the stabilization of activation maps, are further illustrated by visual reconstructions of example inputs (Fig. 6.2c). Initially, at trial 0, the model fails to reproduce the input image, but by trial 100 it generates a blurred yet accurate reconstruction. Similar improvements are observed across T-maze v2 and v3. Notably, when exposed to a novel environment after prior training, the model tends to reuse previously learned reconstruction strategies at the beginning of training (trial 0), suggesting partial transfer or generalization across contexts.

Novelty-modulated hippocampal SWRs promote cognitive map stabilization

Following exposure to novelty, CA1 activity disinhibits VTA dopaminergic neurons, whose feedback to CA1 enhances sharp-wave ripple (SWR) events. We hypothesize that this feedback loop regulates SWR intensity: novel experiences transiently increase SWR occurrence and amplitude, promoting synaptic plasticity and facilitating cognitive-map stabilization. In contrast, during familiar experiences, CA3 generates reliable predictions that match the current sensory input from EC layer III, leading CA1 representations to become spatially constrained, sparser, and more stable. As novelty signals diminish, hippocampal plasticity subsides and SWR activity returns to baseline. Importantly, this modulation is independent of reward, as novelty alone is sufficient to enhance SWRs.

While a direct computational implementation of this SWR-modulated loop remains for future work, our current simulations already provide indirect evidence of its plausibility. As shown in Figs. 6.1 and 6.2, both latent-vector activity and reconstruction dynamics evolve in a synchronized manner across training, indicating a tight coupling between representational stability and overall activation. This relationship suggests that latent-vector activity could serve as a self-regulating signal analogous to hippocampal SWR modulation: as the model becomes familiarized with an environment, activation fields become spatially constrained and overall activity decreases, restoring the functional equivalent of VTA inhibition and returning the system to baseline.

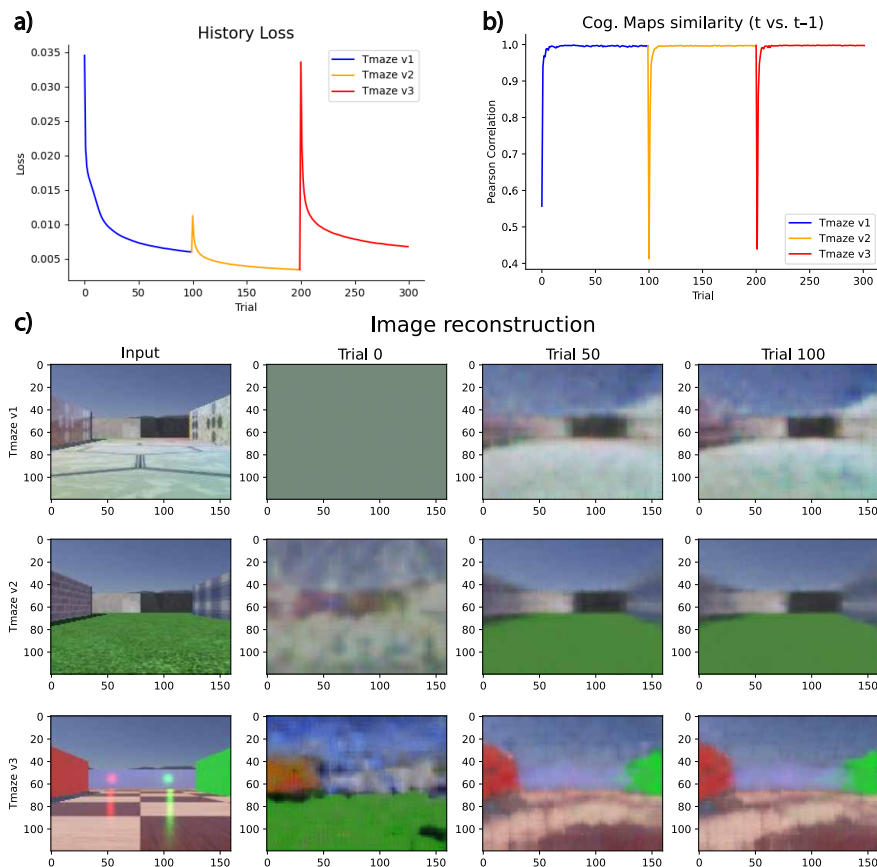


Figure 6.2. Reconstruction performance and cognitive map stabilization in the hippocampal autoencoder. **a)** Reconstruction loss decreases steadily with training within each environment but rises sharply when the dataset is switched to a novel one, reflecting transient instability in representation. **b)** Pearson correlation between activation maps (firing field-like patterns) at trial t and $t + 1$. Exposure to a new environment induces a marked drop in correlation, indicating representational remapping, followed by rapid re-stabilization as training proceeds. **c)** Reconstructions of three example visual inputs (one per T-maze version). When first exposed to a novel environment (trial 0), the model tends to reproduce features of the previously learned environment, suggesting partial transfer of reconstruction strategies across contexts.

While a direct computational implementation of the SWR-modulated loop remains a goal for future work, our current simulations provide indirect evidence supporting its plausibility. As shown in Figs. 6.1 and 6.2, latent-vector activity and reconstruction dynamics evolve in parallel across training, reflecting a coupling between overall activation and representational stability. Within the biological framework, we interpret this relationship as follows: SWRs act as a retraining mechanism that refines hippocampal representations after novelty exposure. As experience accumulates, activation fields become more spatially constrained, signaling that the cognitive map has stabilized. This stabilization would, in turn, restore inhibitory control over VTA dopaminergic neurons, thereby reducing dopamine release to CA1 and returning SWR activity to baseline levels. Thus, the dynamics observed in our model capture the essence of a self-regulating hippocampal-VTA loop, where novelty-driven plasticity naturally transitions into stability once learning is complete.

Motivational signals prime rewarding experience during SWR replays, resulting in their overrepresentation

Once an environment becomes familiar, the introduction of a rewarding stimulus exerts a dual influence on hippocampal function. First, encountering the stimulus generates novelty-related activation through the CA3-CA1 comparator. Second, the consumption of the reward further enhances VTA dopamine release by fulfilling an internal drive. Importantly, these two signals are spatially constrained to the reward site. We hypothesize that their convergence biases the content of hippocampal SWR replays, increasing the probability that experiences near the reward are preferentially reactivated. Over time, this process would strengthen representations of the reward location, producing reward overrepresentation within the hippocampal cognitive map.

To test this hypothesis computationally, we trained our hippocampal autoencoder model in an open-field arena under three successive conditions (Fig. 6.3a): pre-reward bias, reward bias, and post-reward bias. During the first phase (pre-reward bias), the model was trained for 100 epochs using a baseline dataset of 40,000 images collected from a robot performing random navigation. In the second phase (reward bias), the dataset was expanded to 120,000 samples by duplicating images acquired near the reward location (center of arena), thereby biasing the model's learning toward that region (Fig. 6.3b). Finally, in the third phase (post-reward bias), the model was retrained using the

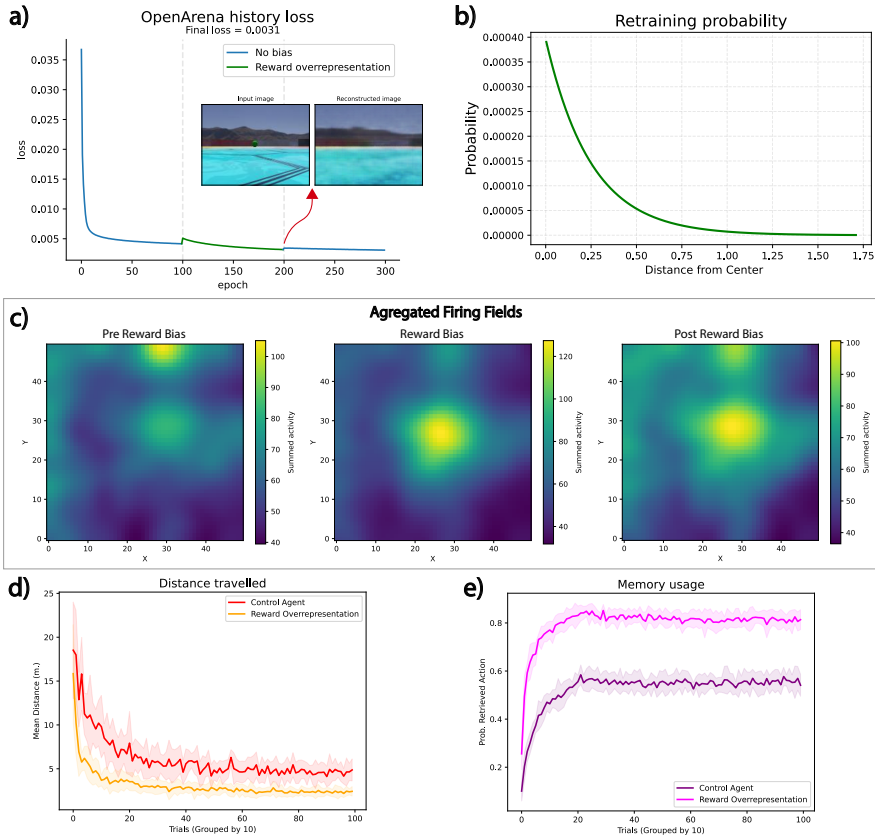


Figure 6.3. Reward-biased replay induces reward overrepresentation and improves goal-directed behavior. **a)** Reconstruction loss across the three training phases (pre-reward bias, reward bias, and post-reward bias), showing progressive improvement in input reconstruction and stable performance across phases. **b)** Probability distribution used to duplicate images during dataset augmentation in the reward bias phase. Samples collected near the reward site (center of the arena) were more frequently duplicated, introducing a spatial learning bias. **c)** Aggregated latent-unit activity at the end of each training phase. Reward-biased training recruited a higher proportion of activation fields near the reward location, producing a clear overrepresentation that partially dissipated once the bias was removed. **d)** Foraging performance measured as the distance travelled before reaching the reward. The agent incorporating the reward-overrepresented model outperformed a control agent, following shorter trajectories to the goal. **e)** Memory usage during the foraging task, quantified as the probability of executing an action retrieved from long-term memory. The reward-overrepresenting agent exhibited greater memory reliance, consistent with enhanced exploitation of prior successful experiences.

original, non-biased dataset to assess whether the induced overrepresentation persisted or dissipated once the reward bias was removed.

At the end of each training phase, we analyzed the activation maps formed by the latent units of the hippocampal autoencoder. By summing all activation maps at each stage, we generated heatmaps reflecting how the model allocated its representational resources across the environment (Fig. 6.3c). During the pre-reward bias phase, activation fields showed moderate clustering near the left and upper boundaries of the arena, as well as around its center. Following the reward bias phase, in which the training data were enriched with samples from the reward region, latent-unit activity became strongly concentrated around the reward site, indicating reward overrepresentation. When the model was subsequently retrained with the non-biased dataset (post-reward bias), this overrepresentation partially dissipated, although residual clustering near the reward location persisted.

These findings demonstrate that the training content can reshape the model's internal representation of space, supporting our hypothesis that reward overrepresentation arises from SWR-mediated replay biased toward reward-related experiences. To evaluate whether such biased cognitive maps provide a functional advantage, we integrated the hippocampal autoencoder with the Sequential Episodic Control (SEC) framework (Freire et al., 2024a; Rosado et al., 2025) and tested the resulting agent in a foraging task.

The agent incorporating the reward-overrepresented model located the reward more efficiently, following shorter trajectories than a control agent equipped with a non-biased hippocampal model (Fig. 6.3d). This improvement in performance was accompanied by a greater reliance on memory (Fig. 6.3e), reflected in an increased tendency to retrieve actions that had been successful in previous trials.

We interpret this enhancement in behavior as a consequence of the fine-grained spatial representations that emerged near the reward site. Such locally dense encoding likely enabled the agent to perform more precise action selections during the final approach to the reward, when navigational demands are highest. In contrast, orienting toward the reward from distant locations depends primarily on coarse spatial cues, which are less sensitive to this bias.

CA3-CA1 mismatch and LHb reward-omission signals drive hippocampal remapping

In familiar environments, reward contingencies can change when a previously rewarding stimulus is removed or relocated. Under these conditions, CA3 continues to generate predictions based on outdated experience, while CA1 receives sensory inputs that no longer match them. Although this mismatch could, in principle, evoke a novelty signal, we hypothesize that it is overridden by inhibitory input from the lateral habenula (LHb) to the VTA, which suppresses dopaminergic activity during reward omission. As a result, hippocampal SWR content would be biased away from replaying reward-associated experiences, progressively weakening their representation within the cognitive map.

This mechanism differs fundamentally from explicit punishment learning. Reward omission does not imply that a specific location should acquire a negative value, since an unrewarded site might yield reward in the future. Avoiding re-exploration would therefore be maladaptive. Instead, omission-driven dopaminergic suppression provides a subtler and more flexible learning signal, gradually fading obsolete reward associations without enforcing aversive avoidance. Punishment circuits may still act in parallel, but omission alone offers a mechanism for adaptive map updating under changing contingencies.

We partially tested this hypothesis by retraining the reward overrepresenting hippocampal autoencoder in a context where the previous reward bias was removed (Fig. 6.3c, right). Under these conditions, the clustering of activation fields around the former reward site gradually dissipated, and the spatial representation became more uniformly distributed. This finding supports the view that reward omission (by suppressing dopaminergic reinforcement of SWRs) is sufficient to drive hippocampal remapping, enabling flexible reorganization of spatial representations when rewards are no longer available.

6.4 Discussion

In this chapter, we have revisited the classical hippocampal-VTA loop and proposed an extended framework that integrates recent insights into hippocampal dynamics, reward modulation, and replay mechanisms. While the original model captured the central role of novelty de-

tection, our revision broadens its scope to include the stabilization of place fields, the dopaminergic regulation of sharp-wave ripples (SWRs), the overrepresentation of reward sites, and the remapping of cognitive maps following reward omission. Together, these mechanisms position the hippocampal-VTA loop as dynamic integrator of spatial, motivational, and novelty-related information that supports adaptive, goal-directed behavior.

To evaluate the plausibility of this framework, we outlined a biologically constrained computational model embodying key hippocampal and dopaminergic interactions. Preliminary computational experiments support that: (i) novelty signals could emerge from diffuse CA1 place fields when predictions are unreliable; (ii) novelty-modulated SWRs could support the stabilization of cognitive maps; (iii) motivationally enhanced novelty signals could be biasing replay toward reward locations, leading to overrepresentation and improved behavioral precision; and (iv) reward omission could suppress dopaminergic reinforcement, inducing hippocampal remapping and restoring sparse representations.

Limitations and Methodological Challenges

Although these results offer proof-of-concept support for the proposed mechanisms, they remain preliminary and reveal several methodological limitations. For instance, when the autoencoder was trained sequentially across environments, the mean spatial coverage of activation fields never returned to the low values reached in the first environment (Fig. 6.1a), and the field topology progressively degraded by the third environment (Fig. 6.1c).

A similar limitation was observed when retraining the reward overrepresenting model with a non-biased dataset. In this case, the overrepresentation around the reward site partially persisted (Fig. 6.3c, right). Further investigations are required to determine whether this effect reflects a methodological limitation, such as insufficient training or the need to explicitly suppress samples near the former reward location (as LHb might do). Alternatively, it may also be the case that this effect represents an implicit memory trace, similar to reinstatement phenomena observed after reward revaluation.

Another potential limitation, not analyzed here, concerns the catastrophic forgetting of previously learned environments when the model

is trained on a novel context. This is a well-known effect in artificial neural networks (Goodfellow et al., 2013) that underscores a key difference between artificial and biological systems.

Theoretical Challenges and Empirical Alignment

Although the presented framework is built upon a broad body of research literature, it also faces theoretical challenges regarding its alignment with physiological evidence. In particular, it has been reported that hippocampal replay is not strictly determined by recent experience (Gupta et al., 2010). For example, during motivational switching tasks, animals often replay trajectories that do not correspond to the currently rewarded experience (e.g., a trajectory to a source of food through the left maze alley when the animal is hungry). Instead, SWR content was relative to the alternative stimulus location (e.g., trajectory toward water source through the right alley under the same motivational condition) (Carey et al., 2019). This observation challenges the assumption that SWR content primarily reflects the most recent or currently relevant reward.

These findings may depend on the temporal dynamics of learning that were not analyzed at a trial-by-trial level. A possible explanation is that replay initially favors motivationally coherent trajectories, facilitating learning. But later, it shifts toward incoherent or alternative representations as part of predictive coding, reward devaluation, or consolidation processes. Indeed, specialized hippocampal cells that encode multiple reward sites simultaneously (Gauthier & Tank, 2018) could account for such flexibility, aligning with our proposal that the hippocampal-VTA loop supports adaptive reorganization rather than fixed encoding.

It also remains an open question whether hippocampal reward overrepresentation also occurs under static reward paradigms, as most empirical evidence comes from dynamic reward settings. We predict that the effect depends on a trade-off between context-related and reward-related novelty signals. In static contexts, reward overrepresentation may fail to emerge because spatially constrained reward signals are masked by the broader novelty of the environment, preventing SWRs from preferentially replaying reward experiences.

Moreover, this mechanism may also encode qualitative aspects of rewards, such as size or palatability, with larger or more salient rewards

producing stronger hippocampal overrepresentation. These predictions could be directly tested in both biological and computational experiments exploring the interaction between novelty, reward magnitude, and replay content.

Implications and Future Directions

Despite these open questions, the extended hippocampal-VTA framework provides a coherent mechanistic account linking novelty detection, reward processing, and replay dynamics. It bridges functional levels from neural signals and plasticity mechanisms to behavioral adaptation, serving as a conceptual bridge between neuroscience and robot autonomy.

Future work should focus on developing a comprehensive cognitive architecture that fully integrates the proposed mechanisms within an embodied agent. Such an implementation would enable the study of emergent behavioral strategies in dynamic environments, providing a more systematic test of the framework. Moreover, coupling this approach with in vivo recordings or neuromodulation studies could validate the predicted relationships between novelty, dopamine, and SWR modulation.

Ultimately, this work illustrates the value of neurorobotic modeling as both a heuristic and experimental tool. By embedding biologically grounded mechanisms into artificial agents, we can formulate testable hypotheses about neural function, refine our understanding of brain-behavior interactions, and guide the design of adaptive systems capable of autonomous learning in complex, changing environments.

7

CHAPTER 7

General discussion

Throughout the preceding chapters, this thesis has examined the computational principles that underpin motivated behavior in the mammalian brain, and how these can be translated into the design of intrinsically motivated artificial agents. By grounding control architectures in biologically constrained models, the work has demonstrated that self-regulation, purposiveness, and adaptive behavior can emerge from the interplay of innate (hardwired) internal needs and external opportunities.

In particular, the contributions span three main domains. First, we showed that our computational models reproduce the reactive orchestration of multiple competing needs, capturing how allostasis can be sustained through attractor dynamics. Second, we demonstrated how hippocampal representations can be modulated by internal states, revealing that motivational cognitive maps can guide efficient goal-directed navigation. Third, we explored how agents can learn from intrinsically rewarding experiences, showing that external rewards only acquire value insofar as they satisfy internal drives.

These findings are supported by a dual approach: a theoretical foundation built upon extensive reviews of neuroscience and cognitive architectures, and an experimental validation where both simulated and real robots were endowed with biologically constrained models. Together, they provide a convergent case for the feasibility of neuro-robotic agents whose adaptive behavior is rooted not in externally defined objectives but in the intrinsic regulation of their own needs.

While each chapter included its own discussion of specific results, this general discussion shifts focus to the broader implications of the work as a whole. Thus, the following sections aim to highlight the potential of biologically constrained models of motivation in informing neuroscience, enhancing robotics, and providing a foundation for social and

ethical dimensions of artificial agents, even while facing a set of limitations.

7.1 Limitations

From Brains to Models: Lost in Translation

Throughout this thesis, the central strategy has been to study the biological brain in search of physiological evidence that can be integrated as computational models of motivation. These models, in turn, provide insight into the computational principles that may underlie brain function and adaptive behavior. The relevance of such models, however, rests on the accuracy with which they reproduce actual data. For this reason, it is important to stress what this thesis is and what it is not. The work presented here should not be interpreted as an attempt to construct a synthetic brain comparable in complexity to a biological one. First, because only a restricted subset of brain functions related to motivated behavior has been addressed. Second, and more importantly, because the biological plausibility of these models remains necessarily limited.

Grace Lindsay has articulated this limitation with clarity: *"Mathematical models are a way to describe a theory about how a biological system works precisely enough to communicate it to others. But biological details inevitably get lost in translation from the real world to the equations. In that sense, all models are wrong the same way all poems are wrong; they capture an essence"* (Lindsay, 2021). This perspective reminds us that every model is both an approximation and a simplification, capturing only fragments of the underlying biological reality.

In this spirit, the models described in this thesis should be viewed as functional abstractions rather than faithful reconstructions. They have proven useful for synthesizing state-of-the-art literature, reproducing previously reported results, and generating new testable predictions. Yet, they operate at a relatively high level of abstraction, focusing on principles of motivational control and their integration within cognitive architectures. Further investigations aiming to continue grounding motivation models in physiology can incorporate details such as laminar organization, neurotransmitter systems, receptor dynamics, and ion-channel communication. Pursuing this line of research would not only refine our explanatory power but also lay the foundations of cognitive

architectures suitable for neuromorphic hardware, where structural and functional constraints of biological circuits can be mirrored more closely.

That is, while the models presented here have captured essential principles of motivation and autonomy, they remain “translations” rather than replicas. Their value lies in the dialogue they enable between theory, physiology, and synthetic implementation, rather than in their fidelity to every biological detail.

Sim-to-Real Gap: The World and its Complexity

A second limitation concerns the well-known gap between robot simulations and real-world deployment. Just as mathematical models of the mind abstract away biological detail, robot simulations abstract away environmental and mechanical complexity. Simulated environments are necessarily simplified: they offer controlled conditions, perfectly reliable sensors and actuators, and the ability to reset trials or respawn objects at will. These affordances are invaluable for rapid experimentation, yet they diverge from the constraints of the physical world. In real settings, experiments are slower, cannot be parallelized or accelerated, and are subject to perturbations and noise that are absent in simulation. Examples include external forces such as wind, the instability and variability of sensory inputs from conventional cameras, or unforeseen events that emerge during interaction with human participants. Such factors complicate both the training and the performance of robotic agents.

Most of the work presented in this thesis has been carried out in simulation, and the performance reported here should be interpreted with this context in mind. Nevertheless, we do not claim that the proposed mechanisms are inapplicable to real robots. On the contrary, much of the modeling concerns internal motivational processes that are not directly dependent on the physical substrate. The challenge is more pronounced in architectures such as the Motivational Hippocampal Autoencoder, which relies heavily on visual data. Here, two sources of complexity emerge. First, the fluctuations of visual features across different conditions (for example, changes in lighting or structural alterations in the environment) pose a greater challenge. However, sensory noise is unlikely to deadly undermine the model's performance since dimensionality reduction filters out much of this variability. Second, such fluctuations may in fact reveal an open physiological question:

does the mammalian hippocampus operate directly on raw sensory input, in which case cognitive maps would vary dramatically between conditions such as day and night? Or does it instead operate on higher-order, feature-based representations (akin to those produced by visual transformers) where contextual parameters such as illumination only modulate a subset of hippocampal neurons? Addressing this question is not only essential for improving robotic architectures but may also provide insights into how biological systems maintain representational stability under variable sensory conditions.

Thus, the sim-to-real gap does not undermine the conceptual validity of the models presented here, but it does highlight the need for future work in embodied real-world settings. Only through such testing can we fully assess the robustness and adaptability of motivational architectures under the complexities of natural environments.

From Prometheus to Humans, from Humans to Humanoids.

A third limitation concerns the suitability of intrinsically motivated architectures for robots that operate in human-centric environments. Contemporary data-driven approaches have so far produced robots with narrow autonomy: they do not independently generate their own goals, and their behavior remains largely constrained by the margins set by their human designers. At first glance, this seems to guarantee safety, but the picture is more complicated. Such robots can still be exploited for malicious purposes, optimized for harmful reward functions, or misinterpret contextual cues due to the limited scope of their design. Moreover, they risk inheriting and amplifying biases present in human training data. Still, as Margaret Boden has pointed out, these machines are not going to take over the world: lacking intrinsic motivations, they do not care about anything, and therefore, they can not aim for power (Boden, 2018).

The situation changes when robots are endowed with intrinsic motivations. Unlike contemporary data-driven systems, intrinsically motivated robots act out of their own regulatory imperatives. They may seek resources, explore environments, or maintain stability independently of explicit human instruction. In principle, this autonomy creates the possibility of competition: an artificial agent might treat humans as rivals for scarce resources, or even as sources of instability to be minimized in the name of self-preservation. While we are still

far from such scenarios, the prospect underscores the importance of critically assessing the risks of introducing artificial motivations into human environments. Despite these concerns, I contend that the study of intrinsically motivated robots is worth pursuing. First, as we saw, it offers a powerful scientific tool for understanding the brain mechanisms by which motivation operates. Second, the limitations of the contemporary data-driven approach suggest that a purely externalist solution is not viable either.

Finally, intrinsically motivated robots (especially those grounded in biologically constrained models) may open the door to the development of moral machines and the emergence of a moral *hybrid society*. In such a society, unethical behaviors may still arise in artificial systems, but unlike in humans, their learning histories are fully transparent. This transparency creates the possibility of analyzing, diagnosing, and designing corrective “educational programs” applicable to both artificial and biological agents. In this sense, intrinsically motivated robots are not only a technological challenge but also an opportunity to re-think how morality is learned, negotiated, and sustained in mixed human-machine communities.

7.2 Understanding the Biological Brain through Testable Predictions

Physiological studies are indispensable for revealing the neural mechanisms behind intelligent behavior, yet they are often costly, technically demanding, and ethically challenging, as they typically require invasive surgery on living animals. In this context, neurorobotics offers a complementary approach: by embedding biologically constrained models into artificial agents, we can test existing hypotheses, generate new ones, and provide predictions that guide future physiological investigations. Synthetic studies, therefore, act as a hypothesis-generating and hypothesis-testing platform, reducing reliance on invasive methods while accelerating the iterative dialogue between theory, experimentation, and evidence.

An illustrative case is provided in Chapter 2, where we developed a neural mass allostatic model inspired by two classes of evidence: (i) the specialization of distinct hypothalamic nuclei for sensing interoceptive states such as hunger, thirst, or temperature, and (ii) inhibitory interactions among these nuclei. From these design principles, we predicted

that the competition between drives would give rise to attractor dynamics capable of autonomously orchestrating motivational priorities, ensuring stability of the self.

Remarkably, this prediction has now been supported by recent physiological studies. Dynamical systems analyses of hypothalamic recordings have revealed approximate line attractors encoding persistent internal states of aggression and sexual receptivity (M. Liu et al., 2024; Nair et al., 2023). More importantly, causal perturbation experiments have demonstrated that hypothalamic circuits intrinsically implement such attractors, exhibiting persistence, integration of inputs, and recovery after off-manifold disturbances (Vinograd et al., 2024). Furthermore, these dynamics are not solely sustained by recurrent connectivity but also critically depend on neuropeptidergic signaling, in particular, oxytocin and vasopressin receptors in VMHvl neurons (Mountoufaris et al., 2024).

These converging results highlight the reciprocal value of synthetic and physiological approaches. While our computational models were inspired by known hypothalamic architecture, they anticipated the existence of attractor dynamics in motivational circuits. A prediction later confirmed by in vivo recordings and perturbation studies. This establishes a powerful methodological loop: candidate mechanisms are modeled and tested in fast, cheap, and safe neurorobotics platforms, physiological studies validate and refine them, and together they advance a mechanistic understanding of how the brain generates intelligent behavior.

7.3 From Core Motivational Systems to Synthetic Awareness

The scientific work reported in this thesis is well aligned with the agenda of the CAVAA (Counterfactual Assessment and Valuation for Awareness Architecture) European project. While the focus of this thesis has been on modeling motivational control as a driver of adaptive behavior, CAVAA extends these ideas toward a broader theory of awareness. Specifically, the project introduces the concept of *virtualization*, understood as the ability to decouple from immediate sensory states and construct an internal model of hidden or counterfactual world states.

Whereas intrinsic motivation provides the foundation for survival by ensuring that agents regulate internal needs and respond adaptively to opportunities, awareness adds the capacity to infer and simulate what cannot be directly perceived. In natural and social environments, many relevant states remain unobservable: the consequences of unexecuted actions, the properties of unexplored regions, or the internal states of conspecifics such as intentions, emotions, and concerns. CAVAA proposes that awareness arises from the virtualization of these hidden states, forming an internal hybrid world that merges perception, memory, and inference into a unified conscious scene.

In this sense, the motivational mechanisms modeled in this thesis can be regarded as core building blocks of a more encompassing awareness architecture. They provide the grounding for survival and purposiveness, upon which virtualization processes can operate to extend the agent's cognitive horizon beyond the here-and-now. CAVAA formalizes this extension through a dedicated virtualization layer, augmenting the classic somatic, reactive, adaptive, and contextual layers of the Distributed Adaptive Control (DAC) architecture (P. F. Verschure, 2012, 2016). This virtualization layer enables parallel prospective simulations, mental time travel, and the construction of counterfactual trajectories, which are then collapsed into a single coherent conscious scene that biases valuation, learning, and decision-making.

Thus, the relation between this thesis and CAVAA is twofold. First, the motivational systems described here supply the “core survival engine” upon which virtualization can build. Without stable motivational control, awareness would lack grounding in the biological imperatives that give it purpose. Second, by embedding biologically constrained motivational models within the broader CAVAA framework, it becomes possible to investigate how awareness contributes to resilience in uncertain environments, the negotiation of social interactions, and the emergence of trust and ethical agency in artificial systems.

Taken together, the step from motivational control to synthetic awareness reflects a natural progression: from ensuring stability of the self, to imagining possible futures, to coexisting with others in shared cognitive and social spaces.

7.4 Scaffolding Abstract Needs on Physiological Foundations

This thesis began by highlighting the centrality of motivation in shaping adaptive behavior. In classical psychology, Maslow's hierarchy of needs offered a powerful metaphor for understanding the progression from basic physiological demands to higher-level aspirations such as self-actualization. Yet, contemporary evidence and theoretical work suggest that this hierarchical view may be too rigid to capture the complexity of real motivational dynamics.

The work presented in this thesis highlights how physiological needs provide a grounding for adaptive behavior, and how higher-order functions such as navigation or learning are modulated by internal states. This resonates with the idea that abstract needs are scaffolded upon physiological ones. Survival imperatives are not left behind when higher goals emerge but remain deeply embedded within them. The case of social thermoregulation, described during the introductory chapter, provided a clear illustration. A social strategy to maintain body temperature within a group is directly tied to survival, yet social rank can be deterministic of which individuals secure the most advantageous positions within the huddle. From this perspective, striving for social dominance or recognition can be understood as unconsciously directed toward increasing social rank and thereby securing better access to resources that ultimately satisfy physiological needs. What looks like the pursuit of abstract goals may, at its root, be anchored in survival regulation.

This framework also casts altruistic behavior in a new light. Helping others, sharing resources, or adhering to social norms can appear selfless on the surface, yet such behaviors enhance group cohesion and stability, which in turn protect the individual. Altruism thus carries a selfish dimension: it functions as a virtualization strategy that secures long-term benefits by ensuring reciprocity, reducing conflict, and maintaining access to the collective resources of the group.

The virtualization framework proposed by the CAVAA project gives these ideas a computational form. Awareness allows agents to anticipate hidden or counterfactual states, such as future scarcity or the intentions of others. In this sense, moral and social behaviors can be interpreted as virtualization-based strategies that enhance resilience across time, while immoral or exploitative behaviors may emerge

when virtualization is misaligned with survival imperatives, privileging short-term gains over long-term stability.

Revisiting Maslow's theory through this lens invites us to move from a static hierarchy to a dynamic, interconnected landscape of needs, where physiological, social, and abstract motivations co-regulate one another. Although preliminary, this perspective underscores a central theme of the thesis: that intelligent, social, and moral dimensions of behavior are not separate from basic survival regulation but continuous with it. Updating Maslow's framework in light of these insights may thus offer a more realistic foundation for both neuroscience and the design of artificial agents.

7.5 Concluding remarks

Investigating the computational principles of motivation and their translation into artificial agents has been both a scientific challenge and a personal journey. This thesis has demonstrated how biologically constrained models can offer a window into the brain's functioning, providing new paths for building purposeful machines. Along the way, I have experienced the excitement of seeing abstract theories come alive in robotic systems and the joy of contributing, even modestly, to the long-standing dialogue between neuroscience and artificial intelligence.

Much remains to be discovered. The questions that motivated this research extend far beyond the scope of this thesis, touching on the very foundations of autonomy, awareness, and social coexistence. Future investigations promise not only to deepen our understanding of the brain but also to inspire new generations of machines that act with resilience, adaptability, and perhaps even moral grounding.

The path forward is demanding but filled with possibilities. If this work has demonstrated anything, it is that bridging biology and robotics does more than advance technology; it enriches our understanding of life itself.

APPENDIX A – Abbreviations

aCell = Adaptive Collaborative Cell
AGV = Automated Guided Vehicle
BCE = Before Current Era
CA1 = Cornu Ammonis 1
CA3 = Cornu Ammonis 3
CAVAA = Counterfactual Assessment and Valuation for Awareness Architecture
CBS = Core Behavioral System
CE = Current Era
Cobot = Collaborative Robot
CRH = Corticotropin-Releasing Hormone
DAC = Distributed Adaptive Control
DAC-HRC = Distributed Adaptive Control for Human-Robot Collaboration
EC = Entorhinal Cortex
FSM = Finite State Machine
GABA = Gamma-Aminobutyric Acid
GDPR = General Data Protection Regulation
GMS = Gain Modulation Score
HFSMs = Hierarchical Finite State Machines
H.M. = Henry Molaison
HPC = Hippocampus
HRL = Homeostatic Reinforcement Learning
HR-RECYCLER = Hybrid Human-Robot RECYcling Plant for ElectriCal and eLEc-tRonic Equipment
HRC = Human-Robot Collaboration
HRI = Human-Robot Interaction
Hz = Hertz
ICTs = Information and Communication Technologies
IM = Interaction Manager
ISO = International Organization for Standardization
LHb = Lateral Habenula
LTM = Long-Term Memory
LTP = Long-Term Potentiation
MDP = Markov Decision Process
MI = Mutual Information
MoHA = Motivational Hippocampal Autoencoder
NAcc = Nucleus Accumbens
NMDAR = N-Methyl-D-Aspartate
OFC = Orbitofrontal Cortex

PFC = Prefrontal Cortex
PPT = Pedunculopontine Nucleus
PV = Parvalbumin
PVH = Paraventricular Hypothalamic Nucleus
RMTg = Rostromedial Tegmental Nucleus
ROS = Robot Operating System
SEC = Sequential Episodic Control
SLAM = Simultaneous Localization and Mapping
SM = State Machines
STM = Short-Term Memory
STT = Somatostatin
Sub = Subiculum
SubT = DARPA's Subterranean Challenge
SWRs = Sharp-Wave Ripples
TP = Task Planner
vStr = Ventral Striatum
VP = Ventral Pallidum
VMHvl = Ventrolateral Subdivision of the Ventromedial Hypothalamus
VTA = Ventral Tegmental Area
WEEE = Waste Electrical and Electronic Equipment

APPENDIX B – Bibliography

- Abdel-Aty, M., & Ding, S. (2024). A matched case-control analysis of autonomous vs human-driven vehicle accidents. *Nature communications*, 15(1), 4931.
- Adams, J. A. (2005). Human-robot interaction design: Understanding user needs and requirements. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49, 447–451.
- Alami, R., Clodic, A., Montreuil, V., Sisbot, E. A., & Chatila, R. (2005). Task planning for human-robot interaction. *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies*, 81–85.
- Alonso-Mora, J., Lohaus, S. H., Leemann, P., Siegwart, R., & Beardsley, P. (2015). Gesture based human-multi-robot swarm interaction and its application to an interactive display. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 5948–5953.
- Ambrose, R. E., Pfeiffer, B. E., & Foster, D. J. (2016). Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron*, 91(5), 1124–1136.
- Amil, A. F., Freire, I. T., & Verschure, P. F. (2024). Discretization of continuous input spaces in the hippocampal autoencoder. *arXiv preprint arXiv:2405.14600*.
- Amil, A. F., & Verschure, P. F. (2021). Supercritical dynamics at the edge-of-chaos underlies optimal decision-making. *Journal of Physics: Complexity*, 2(4), 045017.
- Arkin, R. C. (1998). *Behavior-based robotics*. MIT press.
- Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature*, 543(7647), 719–722.
- Asimov, I. (2004). *I, robot* (Vol. 1). Spectra.
- Axenopoulos, A., Papadopoulos, G., Giakoumis, D., Kostavelis, I., Papadimitriou, A. P., Sillaurren, S., Bastida, L., Oguz, O., Wollherr, D., Garnica, E., et al. (2019). A hybrid human-robot collaborative environment for recycling electrical and electronic equipment. *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBD-Com/IOP/SCI)*, 1754–1759.
- Balcita-Pedicino, J. J., Omelchenko, N., Bell, R., & Sesack, S. R. (2011). The inhibitory influence of the lateral habenula on midbrain dopamine cells: Ultrastructural evidence for indirect mediation via the rostromedial mesopontine tegmental nucleus. *Journal of Comparative Neurology*, 519(6), 1143–1164.

- Baratta, A., Cimino, A., Gnoni, M. G., & Longo, F. (2023). Human robot collaboration in industry 4.0: A literature review. *Procedia Computer Science*, 217, 1887–1895.
- Barbas, H., & Blatt, G. J. (1995). Topographically specific hippocampal projections target functionally distinct prefrontal areas in the rhesus monkey. *Hippocampus*, 5(6), 511–533.
- Barnett, S. A. (2017). *The rat: A study in behavior*. Routledge.
- Barron, A. B., Halina, M., & Klein, C. (2023). Transitions in cognitive evolution. *Proceedings of the Royal Society B*, 290(2002), 20230671.
- Basu, R., Gebauer, R., Herfurth, T., Kolb, S., Golipour, Z., Tchumatchenko, T., & Ito, H. T. (2021). The orbitofrontal cortex maps future navigational goals. *Nature*, 599(7885), 449–452.
- Bauer, A., Wollherr, D., & Buss, M. (2008). Human–robot collaboration: A survey. *International Journal of Humanoid Robotics*, 5(01), 47–66.
- Benna, M. K., & Fusi, S. (2021). Place cells may simply be memory cells: Memory compression leads to spatial tuning and history dependence. *Proceedings of the National Academy of Sciences*, 118(51), e2018422118.
- Bernard, C. (1865). *Introduction à l'étude de la médecine expérimentale*. JB Baillière.
- Best, G., & Moghadam, P. (2014). An evaluation of multi-modal user interface elements for tablet-based robot teleoperation. *Proc. of ARAA ACRA*.
- Binmore, K. (2005). *Natural justice*. Oxford university press.
- Blouet, C., & Schwartz, G. J. (2010). Hypothalamic nutrient sensing in the control of energy homeostasis. *Behavioural brain research*, 209(1), 1–12.
- Boden, M. (2018, June). The future of intelligence: What matters? — the margaret boden lecture, lecture one [Inaugural Margaret Boden Lecture, filmed at Cambridge Union Society, 5 June 2018]. <https://www.youtube.com/watch?v=zNr2BSzqObY&t=2048s>
- Bohren, J., & Cousins, S. (2010). The smach high-level executive [ros news]. *IEEE Robotics & Automation Magazine*, 17(4), 18–20.
- Bos, J. J., Vinck, M., Marchesi, P., Keestra, A., van Mourik-Donga, L. A., Jackson, J. C., Verschure, P. F., & Pennartz, C. M. (2019). Multiplexing of information about self and others in hippocampal ensembles. *Cell Reports*, 29(12), 3859–3871.
- Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology*. MIT press.
- Brooks, R. (2003). A robust layered control system for a mobile robot. *IEEE journal on robotics and automation*, 2(1), 14–23.
- Bruynseels, K., Santoni de Sio, F., & Van den Hoven, J. (2018). Digital twins in health care: Ethical implications of an emerging engineering paradigm. *Frontiers in genetics*, 9, 31.

- Buck, C. L., & Barnes, B. M. (1999). Annual cycle of body composition and hibernation in free-living arctic ground squirrels. *Journal of Mammalogy*, 80(2), 430–442.
- Burnett, C. J., Li, C., Webber, E., Tsaousidou, E., Xue, S. Y., Brüning, J. C., & Krashes, M. J. (2016). Hunger-driven motivational state competition. *Neuron*, 92(1), 187–201.
- Buzsáki, G., & Tingley, D. (2023). Cognition from the body-brain partnership: Exaptation of memory. *Annual review of neuroscience*, 46(1), 191–210.
- Cabanac, M. (1971). Physiological role of pleasure: A stimulus can feel pleasant or unpleasant depending upon its usefulness as determined by internal signals. *Science*, 173(4002), 1103–1107.
- Campbell, L. A., Tkaczynski, P. J., Lehmann, J., Mouna, M., & Majolo, B. (2018). Social thermoregulation as a potential mechanism linking sociality and fitness: Barbary macaques with more social partners form larger huddles. *Scientific reports*, 8(1), 6074.
- Canals, M., & Bozinovic, F. (2011). Huddling behavior as critical phase transition triggered by low temperatures. *Complexity*, 17(1), 35–43.
- Cannon, W. B. (1939). *The wisdom of the body*. Norton & Co.
- Carey, A. A., Tanaka, Y., & van Der Meer, M. A. (2019). Reward revaluation biases hippocampal replay content away from the preferred outcome. *Nature neuroscience*, 22(9), 1450–1459.
- Castro, A., Silva, F., & Santos, V. (2021). Trends of Human-Robot Collaboration in Industry Contexts: Handover, Learning, and Metrics. *Sensors*, 21(12), 4113. <https://doi.org/10.3390/s21124113>
- Cette, G., Devillard, A., & Spiezia, V. (2021). The contribution of robots to productivity growth in 30 oecd countries over 1975–2019. *Economics Letters*, 200, 109762.
- Charalambous, G., Fletcher, S., & Webb, P. (2015). Identifying the key organisational human factors for introducing human-robot collaboration in industry: An exploratory study. *The International Journal of Advanced Manufacturing Technology*, 81(9), 2143–2155. <https://doi.org/10.1007/s00170-015-7335-4>
- Christoforakos, L., Gallucci, A., Surmava-Große, T., Ullrich, D., & Diefenbach, S. (2021). Can robots earn our trust the same way humans do? a systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in hri. *Frontiers in Robotics and AI*, 8, 79.
- Chrobak, J., & Napier, T. (1993). Opioid and gaba modulation of accumbens-evoked ventral pallidal activity. *Journal of Neural Transmission/General Section JNT*, 93, 123–143.

- Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1111/1467-8284.00096>
- Connell, J. H. (1989). *A colony architecture for an artificial creature* (tech. rep.).
- Cos, I., Canamero, L., Hayes, G. M., & Gillies, A. (2013). Hedonic value: Enhancing adaptation for motivated agents. *Adaptive Behavior*, 21(6), 465–483.
- Croatti, A., Gabellini, M., Montagna, S., & Ricci, A. (2020). On the integration of agents and digital twins in healthcare. *Journal of Medical Systems*, 44, 1–8.
- Diba, K., & Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature neuroscience*, 10(10), 1241–1242.
- Dictionary, M.-W. (2002). Merriam-webster. On-line at <http://www.mw.com/home.htm>, 8(2), 23.
- Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J., Zhou, M., & Hon, H.-W. (2019). Unified language model pre-training for natural language understanding and generation. *Advances in neural information processing systems*, 32.
- Dor, D. (2023). Communication for collaborative computation: Two major transitions in human evolution. *Philosophical Transactions of the Royal Society B*, 378(1872), 20210404.
- Douglas, R. J. (1967). The hippocampus and behavior. *Psychological bulletin*, 67(6), 416.
- Doyle, J. C., & Csete, M. (2011). Architecture, constraints, and behavior. *Proceedings of the National Academy of Sciences*, 108(supplement_3), 15624–15630.
- Dupret, D., O’neill, J., Pleydell-Bouverie, B., & Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature neuroscience*, 13(8), 995–1002.
- Dynamics, B. (2021, August). Atlas | partners in parkour [YouTube video]. <https://www.youtube.com/watch?v=tF4DML7FIWk>
- Edmonds, M., Gao, F., Liu, H., Xie, X., Qi, S., Rothrock, B., Zhu, Y., Wu, Y. N., Lu, H., & Zhu, S.-C. (2019). A tale of two explanations: Enhancing human trust by explaining robot behavior. *Science Robotics*, 4(37).
- Eichenbaum, H. (2017). The role of the hippocampus in navigation is memory. *Journal of neurophysiology*, 117(4), 1785–1796.
- Farrell, K., Lak, A., & Saleem, A. B. (2022). Midbrain dopamine neurons signal phasic and ramping reward prediction error during goal-directed navigation. *Cell Reports*, 41(2).
- Fibla, M. S., Bernardet, U., & Verschure, P. F. (2010). Allostatic control for robot behaviour regulation: An extension to path planning. *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1935–1942.

- Finch, D. M. (1996). Neurophysiology of converging synaptic inputs from the rat prefrontal cortex, amygdala, midline thalamus, and hippocampal formation onto single neurons of the caudate/putamen and nucleus accumbens. *Hippocampus*, 6(5), 495–512.
- Fischer, T., Puigbò, J.-Y., Camilleri, D., Nguyen, P. D., Moulin-Frier, C., Lallée, S., Metta, G., Prescott, T. J., Demiris, Y., & Verschure, P. F. (2018). Icube-hri: A software framework for complex human–robot interaction scenarios on the icub humanoid robot. *Frontiers in Robotics and AI*, 5, 22.
- Floresco, S. B., Todd, C. L., & Grace, A. A. (2001). Glutamatergic afferents from the hippocampus to the nucleus accumbens regulate activity of ventral tegmental area dopamine neurons. *Journal of Neuroscience*, 21(13), 4915–4922.
- Floresco, S. B., West, A. R., Ash, B., Moore, H., & Grace, A. A. (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nature neuroscience*, 6(9), 968–973.
- Foukarakis, M., Leonidis, A., Antona, M., & Stephanidis, C. (2014). Combining finite state machine and decision-making tools for adaptable robot behavior. *Universal Access in Human-Computer Interaction. Aging and Assistive Environments: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22–27, 2014, Proceedings, Part III* 8, 625–635.
- Freire, I. T., Amil, A. F., & Verschure, P. F. (2024a). Sequential memory improves sample and memory efficiency in episodic control. *Nature Machine Intelligence*, 1–13.
- Freire, I. T., Arsiwalla, X. D., Puigbò, J.-Y., & Verschure, P. (2023). Modeling theory of mind in dyadic games using adaptive feedback control. *Information*, 14(8), 441.
- Freire, I. T., Guerrero-Rosado, O., Amil, A. F., & Verschure, P. F. (2024b). Socially adaptive cognitive architecture for human-robot collaboration in industrial settings. *Frontiers in Robotics and AI*, 11, 1248646.
- Freire, I. T., Moulin-Frier, C., Sanchez-Fibla, M., Arsiwalla, X. D., & Verschure, P. F. (2020a). Modeling the formation of social conventions from embodied real-time interactions. *PloS one*, 15(6), e0234434.
- Freire, I. T., Moulin-Frier, C., Sanchez-Fibla, M., Arsiwalla, X. D., & Verschure, P. F. (2020b). Modeling the formation of social conventions from embodied real-time interactions. *PloS one*, 15(6), e0234434.
- Freire, I. T., Urikh, D., Arsiwalla, X. D., & Verschure, P. F. (2020c). Machine morality: From harm-avoidance to human-robot cooperation. *Biomimetic and Biohybrid Systems: 9th International Conference, Living Machines 2020, Freiburg, Germany, July 28–30, 2020, Proceedings* 9, 116–127.

- French, S. J., & Totterdell, S. (2002). Hippocampal and prefrontal cortical inputs monosynaptically converge with individual projection neurons of the nucleus accumbens. *Journal of Comparative Neurology*, 446(2), 151–165.
- Frey, U., Matthies, H., Reymann, K. G., & Matthies, H. (1991). The effect of dopaminergic d1 receptor blockade during tetanization on the expression of long-term potentiation in the rat ca1 region in vitro. *Neuroscience letters*, 129(1), 111–114.
- Frey, U., Schroeder, H., et al. (1990). Dopaminergic antagonists prevent long-term maintenance of posttetanic ltp in the ca1 region of rat hippocampal slices. *Brain research*, 522(1), 69–75.
- Fujisawa, S., & Buzsáki, G. (2011). A 4 hz oscillation adaptively synchronizes prefrontal, vta, and hippocampal activities. *Neuron*, 72(1), 153–165.
- Fyhn, M., Molden, S., Hollup, S., Moser, M.-B., & Moser, E. I. (2002). Hippocampal neurons responding to first-time dislocation of a target object. *Neuron*, 35(3), 555–566.
- Gasbarri, A., Sulli, A., & Packard, M. G. (1997). The dopaminergic mesencephalic projections to the hippocampal formation in the rat. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 21(1), 1–22.
- Gauthier, J. L., & Tank, D. W. (2018). A dedicated population for reward coding in the hippocampus. *Neuron*, 99(1), 179–193.
- Gomperts, S. N., Kloosterman, F., & Wilson, M. A. (2015). Vta neurons coordinate with the hippocampal reactivation of spatial experience. *Elife*, 4, e05360.
- Goodfellow, I. J., Mirza, M., Xiao, D., Courville, A., & Bengio, Y. (2013). An empirical investigation of catastrophic forgetting in gradient-based neural networks. *arXiv preprint arXiv:1312.6211*.
- Goodrich, M. A., & Olsen, D. R. (2003). Seven principles of efficient human robot interaction. *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*, 4, 3942–3948.
- Gould, T. D., Dao, D. T., & Kovacsics, C. E. (2009). The open field test. *Mood and anxiety related phenotypes in mice*, 1–20.
- Grillo, F. (2019). *Hero of alexandria's automata: A critical edition and translation, including a commentary on book one* [Doctoral dissertation, University of Glasgow].
- Grove, J. C., Gray, L. A., La Santa Medina, N., Sivakumar, N., Ahn, J. S., Corpuz, T. V., Berke, J. D., Kreitzer, A. C., & Knight, Z. A. (2022). Dopamine subsystems that track internal states. *Nature*, 608(7922), 374–380.
- Guerrero Rosado, O., Amil, A. F., Freire, I. T., Vinck, M., & Verschure, P. F. (2025). Biomimetic self-regulation in intrinsically motivated robots. *Artificial Life Conference Proceedings 37*, 2025(1), 49.

- Guerrero-Rosado, O., Amil, A. F., Freire, I. T., & Verschure, P. F. (2022). Drive competition underlies effective allostatic orchestration. *Frontiers in Robotics and AI*, 9, 1052998.
- Guerrero-Rosado, O., & Verschure, P. F. (2020). Distributed adaptive control: An ideal cognitive architecture candidate for managing a robotic recycling plant. *Conference on Biomimetic and Biohybrid Systems*, 153–164.
- Gupta, A. S., Van Der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, 65(5), 695–705.
- Hawkins, R. X. D., & Goldstone, R. L. (2016). The Formation of Social Conventions in Real-Time Environments (C. T. Bauch, Ed.). *PLOS ONE*, 11(3), e0151670. <https://doi.org/10.1371/journal.pone.0151670>
- Hollup, S. A., Molden, S., Donnett, J. G., Moser, M.-B., & Moser, E. I. (2001). Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *Journal of Neuroscience*, 21(5), 1635–1644.
- Houston, A., & McFarland, D. (1976). On the measurement of motivational variables. *Animal Behaviour*, 24(2), 459–475.
- Hu, H., Cui, Y., & Yang, Y. (2020). Circuits and functions of the lateral habenula in health and in disease. *Nature Reviews Neuroscience*, 21(5), 277–295.
- Huang, Y.-Y., & Kandel, E. R. (1995). D1/d5 receptor agonists induce a protein synthesis-dependent late potentiation in the ca1 region of the hippocampus. *Proceedings of the National Academy of Sciences*, 92(7), 2446–2450.
- Hull, C. L. (1945). Principles of behavior.
- Hull, C. L. (1952). A behavior system; an introduction to behavior theory concerning the individual organism.
- Humza, R., Scholz, O., Mokhtar, M., Timmis, J., & Tyrrell, A. (2009). Towards energy homeostasis in an autonomous self-reconfigurable modular robotic organism. *2009 Computation World: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns*, 21–26.
- Ieropoulos, I., Greenman, J., & Melhuish, C. (2003). Imitating metabolism: Energy autonomy in biologically inspired robots.
- Ihalainen, J., Riekkinen Jr, P., & Feenstra, M. (1999). Comparison of dopamine and noradrenaline release in mouse prefrontal cortex, striatum and hippocampus using microdialysis. *Neuroscience letters*, 277(2), 71–74.
- Inkulu, A. K., Bahubalendruni, M. R., Dara, A., & K., S. (2021). Challenges and opportunities in human robot collaboration context of Industry 4.0 - a state of the art review [Publisher: Emerald Publishing Limited]. *Industrial Robot: the international journal of robotics research and application*, 49(2), 226–239. <https://doi.org/10.1108/IR-04-2021-0077>

- ISO. (2011). ISO 10218-1:2011 Robots and robotic devices—Safety requirements for industrial robots—Part 1: Robots.
- ISO. (2016). ISO/TS 15066:2016 Robots and robotic devices — Collaborative robots.
- Jhou, T. C., Fields, H. L., Baxter, M. G., Saper, C. B., & Holland, P. C. (2009). The rostromedial tegmental nucleus (rmtg), a gabaergic afferent to mid-brain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron*, 61(5), 786–800.
- Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience*, 10(1), 100–107.
- Jimenez-Rodriguez, A., Prescott, T. J., Schmidt, R., & Wilson, S. (2020). A framework for resolving motivational conflict via attractor dynamics. *Conference on Biomimetic and Biohybrid Systems*, 192–203.
- Johannsmeier, L., & Haddadin, S. (2016). A hierarchical human-robot interaction-planning framework for task allocation in collaborative industrial assembly processes. *IEEE Robotics and Automation Letters*, 2(1), 41–48.
- Jokinen, K., & Wilcock, G. (2015). Modelling user experience in human-robot interactions. *Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction: Second International Workshop, MA3HMI 2014, Held in Conjunction with INTERSPEECH 2014, Singapore, Singapore, September 14, 2014, Revised Selected Papers 2*, 45–56.
- Kawaguchi, Y. (1997). Selective cholinergic modulation of cortical gabaergic cell subtypes. *Journal of neurophysiology*, 78(3), 1743–1747.
- Kennedy, P. J., & Shapiro, M. L. (2009). Motivational states activate distinct hippocampal representations to guide goal-directed behaviors. *Proceedings of the National Academy of Sciences*, 106(26), 10805–10810.
- Keramati, M., & Gutkin, B. (2014). Homeostatic reinforcement learning for integrating reward collection and physiological stability. *Elife*, 3.
- Kernbach, S., Meister, E., Schlachter, F., Jebens, K., Szymanski, M., Liedke, J., Laneri, D., Winkler, L., Schmickl, T., Thenius, R., et al. (2008). Symbiotic robot organisms: Replicator and symbion projects. *Proceedings of the 8th workshop on performance metrics for intelligent systems*, 62–69.
- Kim, S., Jung, D., & Royer, S. (2020). Place cell maps slowly develop via competitive learning and conjunctive coding in the dentate gyrus. *Nature communications*, 11(1), 4550.
- Kleinman, M. R., & Foster, D. J. (2024). Spatial localization of hippocampal replay requires dopamine signaling. *bioRxiv*, 2024–06.
- Kloza, D., Van Dijk, N., Casiraghi, S., Vazquez Maymir, S., Roda, S., Tanas, A., & Konstantinou, I. (2019). Towards a method for data protection impact

- assessment: Making sense of gdpr requirements. *d. pia. lab Policy Brief*, 1, 1–8.
- Kobayashi, Y., Inoue, Y., Yamamoto, M., Isa, T., & Aizawa, H. (2002). Contribution of pedunculopontine tegmental nucleus neurons to performance of visually guided saccade tasks in monkeys. *Journal of neurophysiology*, 88(2), 715–731.
- Kok, B. C., & Soh, H. (2020). Trust in robots: Challenges and opportunities. *Current Robotics Reports*, 1–13.
- Kontogianni, A., Kabassi, K., & Alepis, E. (2018). Designing a smart tourism mobile application: User modelling through social networks' user implicit data. *Social Informatics: 10th International Conference, SocInfo 2018, St. Petersburg, Russia, September 25-28, 2018, Proceedings, Part II 10*, 148–158.
- Koyama, Y., Jodo, E., & Kayama, Y. (1994). Sensory responsiveness of “broad-spike” neurons in the laterodorsal tegmental nucleus, locus coeruleus and dorsal raphe of awake rats: Implications for cholinergic and monoaminergic neuron-specific responses. *Neuroscience*, 63(4), 1021–1031.
- Krishnan, S., Heer, C., Cherian, C., & Sheffield, M. E. (2022). Reward expectation extinction restructures and degrades ca1 spatial maps through loss of a dopaminergic reward proximity signal. *Nature communications*, 13(1), 6662.
- Kumar, A., Braud, T., Tarkoma, S., & Hui, P. (2020). Trustworthy ai in the age of pervasive computing and big data. *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 1–6.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, 40.
- Lallée, S., Vouloutsis, V., Munoz, M. B., Grechuta, K., Llobet, J.-Y. P., Sarda, M., & Verschure, P. F. (2015). Towards the synthetic self: Making others perceive me as an other. *Paladyn, Journal of Behavioral Robotics*, 6(1).
- Larkin, M. C., Lykken, C., Tye, L. D., Wickelgren, J. G., & Frank, L. M. (2014). Hippocampal output area ca1 broadcasts a generalized novelty signal during an object-place recognition task. *Hippocampus*, 24(7), 773–783.
- Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & information systems engineering*, 6, 239–242.
- Laurençon, H., Ségerie, C.-R., Lussange, J., & Gutkin, B. S. (2021). Continuous homeostatic reinforcement learning for self-regulated autonomous agents. *arXiv preprint arXiv:2109.06580*.

- Lee, C. R., Chen, A., & Tye, K. M. (2021). The neural circuitry of social homeostasis: Consequences of acute versus chronic social isolation. *Cell*, 184(6), 1500–1516.
- Legault, M., Rompré, P.-P., & Wise, R. A. (2000). Chemical stimulation of the ventral hippocampus elevates nucleus accumbens dopamine by activating dopaminergic neurons of the ventral tegmental area. *Journal of Neuroscience*, 20(4), 1635–1642.
- Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., & Moser, M.-B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, 309(5734), 619–623.
- Lever, C., Burton, S., Jeewajee, A., Wills, T. J., Cacucci, F., Burgess, N., & O'Keefe, J. (2010). Environmental novelty elicits a later theta phase of firing in ca1 but not subiculum. *Hippocampus*, 20(2), 229–234.
- Levin, M. (2021). Life, death, and self: Fundamental questions of primitive cognition viewed through the lens of body plasticity and synthetic organisms. *Biochemical and Biophysical Research Communications*, 564, 114–133.
- Levinson, S. C. (2020). On the human" interaction engine". In *Roots of human sociality* (pp. 39–69). Routledge.
- Li, M., Han, Y., Aburn, M. J., Breakspear, M., Poldrack, R. A., Shine, J. M., & Lizzier, J. T. (2019). Transitions in information processing dynamics at the whole-brain network level are driven by alterations in neural gain. *PLoS computational biology*, 15(10), e1006957.
- Lindsay, G. (2021). *Models of the mind: How physics, engineering and mathematics have shaped our understanding of the brain*. Bloomsbury Publishing.
- Lisman, J. E. (1999). Relating hippocampal circuitry to function: Recall of memory sequences by reciprocal dentate–ca3 interactions. *Neuron*, 22(2), 233–242.
- Lisman, J. E., & Grace, A. A. (2005). The hippocampal-vta loop: Controlling the entry of information into long-term memory. *Neuron*, 46(5), 703–713.
- Liu, H., & Wang, L. (2018). Gesture recognition for human-robot collaboration: A review. *International Journal of Industrial Ergonomics*, 68, 355–367.
- Liu, M., Nair, A., Coria, N., Linderman, S. W., & Anderson, D. J. (2024). Encoding of female mating dynamics by a hypothalamic line attractor. *Nature*, 634(8035), 901–909.
- Lodge, D. J., & Grace, A. A. (2006). The hippocampus modulates dopamine neuron responsivity by regulating the intensity of phasic neuron activation. *Neuropsychopharmacology*, 31(7), 1356–1361.
- Lollini, M. (2012). On becoming human: The verum factum principle and giambattista vico's humanism. *MIn*, 127(1), S21–S31.

- Lu, Y., & Metz, C. (2023). Cruise's driverless service suspended in san francisco [Accessed: 2025-09-16]. *The New York Times*. <https://www.nytimes.com/2023/10/24/technology/cruise-driverless-san-francisco-suspended.html>
- Luna-Reyes, L. F., & Andersen, D. L. (2003). Collecting and analyzing qualitative data for system dynamics: Methods and models. *System Dynamics Review: The Journal of the System Dynamics Society*, 19(4), 271–296.
- Luz, R., Corujeira, J., Grisoni, L., Giraud, F., Silva, J. L., & Ventura, R. (2019). On the use of haptic tablets for ugv teleoperation in unstructured environments: System design and evaluation. *IEEE Access*, 7, 95443–95454.
- Ma, Z., Turrigiano, G. G., Wessel, R., & Hengen, K. B. (2019). Cortical circuit dynamics are homeostatically tuned to criticality in vivo. *Neuron*, 104(4), 655–664.
- Maffei, G., Santos-Pata, D., Marcos, E., Sánchez-Fibla, M., & Verschure, P. F. (2015). An embodied biologically constrained model of foraging: From classical and operant conditioning to adaptive real-world behavior in dac-x. *Neural Networks*, 72(November), 88–108. <https://doi.org/10.1016/j.neunet.2015.10.004>
- Malik, A. A., & Brem, A. (2021). Digital twins for collaborative robots: A case study in human-robot interaction. *Robotics and Computer-Integrated Manufacturing*, 68, 102092.
- Marshall, J. A., Favreau-Peigné, A., Fromhage, L., Mcnamara, J. M., Meah, L. F., & Houston, A. I. (2015). Cross inhibition improves activity selection when switching incurs time costs. *Current zoology*, 61(2), 242–250.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological review*, 50(4), 370.
- Matsumoto, M., & Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, 447(7148), 1111–1115.
- Maturana, H., et al. (1973). *De maquinas y seres vivos: Una teoría sobre la organización biológica*. Editorial Universitaria.
- Maturana, H. R., & Varela, F. J. (2012). *Autopoiesis and cognition: The realization of the living* (Vol. 42). Springer Science & Business Media.
- Mayor, A. (2018). *Gods and robots: Myths, machines, and ancient dreams of technology*. Princeton University Press.
- Mazachek, K. (2020). Robots and the economy: The role of automation in productivity growth. <https://blog.trade.gov/2020/12/22/robots-and-the-economy-the-role-of-automation-in-productivity-growth/>
- McNamara, C. G., Tejero-Cantero, Á., Trouche, S., Campo-Urriza, N., & Dupret, D. (2014). Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nature neuroscience*, 17(12), 1658–1660.

- Melhuish, C., Ieropoulos, I., Greenman, J., & Horsfield, I. (2006). Energetically autonomous robots: Food for thought. *Autonomous Robots*, 21(3), 187–198.
- Merker, B. (2013). The efference cascade, consciousness, and its self: Naturalizing the first person pivot of action control. *Frontiers in Psychology*, 4, 501.
- Michel, O. (2004). Cyberbotics Ltd. webots™: Professional mobile robot simulation. *International Journal of Advanced Robotic Systems*, 1(1), 5.
- Miyawaki, T., Norimoto, H., Ishikawa, T., Watanabe, Y., Matsuki, N., & Ikegaya, Y. (2014). Dopamine receptor activation reorganizes neuronal ensembles during hippocampal sharp waves in vitro. *PloS one*, 9(8), e104438.
- Mizumori, S. J., & Baker, P. M. (2017). The lateral habenula and adaptive behaviors. *Trends in neurosciences*, 40(8), 481–493.
- Mogenson, G. J., Brudzynski, S. M., Wu, M., Yang, C. R., & Yim, C. C. (2019). From motivation to action: A review of dopaminergic regulation of limbic-nucleus accumbens-ventral pallidum-pedunculo-pontine nucleus circuitries involved in limbic-motor integration. *Limbic motor circuits and neuropsychiatry*, 193–236.
- Montévil, M., & Mossio, M. (2015). Biological organisation as closure of constraints. *Journal of theoretical biology*, 372, 179–191.
- Mossio, M., & Moreno, A. (2015). Biological autonomy: A philosophical and theoretical enquiry. history, philosophy, and theory of life sciences series.
- Moulin-Frier, C., Fischer, T., Petit, M., Pointeau, G., Puigbo, J.-Y., Pattacini, U., Low, S. C., Camilleri, D., Nguyen, P., Hoffmann, M., et al. (2017). Dac-h3: A proactive robot cognitive architecture to acquire and express knowledge about the world and the self. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4), 1005–1022.
- Mountoufaris, G., Nair, A., Yang, B., Kim, D.-W., Vinograd, A., Kim, S., Linderman, S. W., & Anderson, D. J. (2024). A line attractor encoding a persistent internal state requires neuropeptide signaling. *Cell*, 187(21), 5998–6015.
- Musée d'art et d'histoire Neuchâtel. (2023). Les automates Jaquet-Droz [[Online; accessed 3-October-2024]].
- Nair, A., Karigo, T., Yang, B., Ganguli, S., Schnitzer, M. J., Linderman, S. W., Anderson, D. J., & Kennedy, A. (2023). An approximate line attractor in the hypothalamus encodes an aggressive state. *Cell*, 186(1), 178–193.
- Nakamura, K. (2011). Central circuitries for body temperature regulation and fever. *American journal of Physiology-Regulatory, integrative and comparative Physiology*, 301(5), R1207–R1228.
- Needham, J. (1974). *Science and civilisation in china* (Vol. 5). Cambridge University Press.
- Negri, E., Fumagalli, L., & Macchi, M. (2017). A review of the roles of digital twin in cps-based production systems. *Procedia manufacturing*, 11, 939–948.

- Newman, W. R. (1998, September). Alchemy, Domination, and Gender. In *A House Built on Sand: Exposing Postmodernist Myths About Science*. Oxford University Press.
- Nieh, E. H., Vander Weele, C. M., Matthews, G. A., Presbrey, K. N., Wichmann, R., Leppla, C. A., Izadmehr, E. M., & Tye, K. M. (2016). Inhibitory input from the lateral hypothalamus to the ventral tegmental area disinhibits dopamine neurons and promotes behavioral activation. *Neuron*, 90(6), 1286–1298.
- Nocks, L. (1998). The golem: Between the technological and the divine. *Journal of Social and Evolutionary Systems*, 21(3), 281–303.
- O'Donnell, P., & Grace, A. (1998). Phencyclidine interferes with the hippocampal gating of nucleus accumbens neuronal activity in vivo. *Neuroscience*, 87(4), 823–830.
- O'Donnell, P., & Grace, A. A. (1995). Synaptic interactions among excitatory afferents to nucleus accumbens neurons: Hippocampal gating of prefrontal cortical input. *Journal of Neuroscience*, 15(5), 3622–3639.
- O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental neurology*, 51(1), 78–109.
- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*.
- O'Madagain, C., & Tomasello, M. (2022). Shared intentionality, reason-giving and the evolution of human culture. *Philosophical Transactions of the Royal Society B*, 377(1843), 20200320.
- Osterhout, J. A., Kapoor, V., Eichhorn, S. W., Vaughn, E., Moore, J. D., Liu, D., Lee, D., DeNardo, L. A., Luo, L., Zhuang, X., et al. (2022). A preoptic neuronal population controls fever and appetite during sickness. *Nature*, 606(7916), 937–944.
- Papies, E. K., Aarts, H., et al. (2016). Automatic self-regulation: From habit to goal pursuit: Handbook of self regulation: Research, theory, and applications.
- Pavlov, I. P. (1955). *Selected works* (Vol. 1). Foreign Languages Publishing House.
- Peral, M., Sanfeliu, A., & Garrell, A. (2022). Efficient hand gesture recognition for human-robot interaction. *IEEE Robotics and Automation Letters*, 7(4), 10272–10279.
- Perez, E., Strub, F., De Vries, H., Dumoulin, V., & Courville, A. (2018). Film: Visual reasoning with a general conditioning layer. *Proceedings of the AAAI conference on artificial intelligence*, 32(1).
- Petzold, A., Van den Munkhof, H. E., Figge-Schlensook, R., & Korotkova, T. (2023). Complementary lateral hypothalamic populations resist hunger pressure to balance nutritional and social needs. *Cell Metabolism*, 35(3), 456–471.

- Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., & Castelfranchi, C. (2019). The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of life reviews*, 28, 1–21.
- Pradalier, C. (2017). *A task scheduler for ros* [Doctoral dissertation, UMI 2958 GeorgiaTech-CNRS].
- Qian, S., Yan, S., Pang, R., Zhang, J., Liu, K., Shi, Z., Wang, Z., Chen, P., Zhang, Y., Luo, T., et al. (2022). A temperature-regulated circuit for feeding behavior. *Nature Communications*, 13(1), 1–17.
- Raessa, M., Chen, J. C. Y., Wan, W., & Harada, K. (2020). Human-in-the-loop robotic manipulation planning for collaborative assembly. *IEEE Transactions on Automation Science and Engineering*, 17(4), 1800–1813.
- Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: A review. *Medical and biological engineering and computing*, 44, 1031–1051.
- Ranck Jr, J. B. (1973). Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats: Part i. behavioral correlates and firing repertoires. *Experimental neurology*, 41(2), 462–531.
- Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies*, 1(1), 33–36.
- Robla-Gómez, S., Becerra, V. M., Llata, J. R., González-Sarabia, E., Torre-Ferrero, C., & Pérez-Oria, J. (2017). Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments [Conference Name: IEEE Access]. *IEEE Access*, 5, 26754–26773. <https://doi.org/10.1109/ACCESS.2017.2773127>
- Rosado, O. G., Amil, A. F., Freire, I. T., Vinck, M., & Verschure, P. F. (2025). Motivational cognitive maps allow robot biomimetic autonomy. *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 14402–14408.
- Rosenberg-Kima, R., Koren, Y., Yachini, M., & Gordon, G. (2019). Human-robot-collaboration (hrc): Social robots as teaching assistants for training activities in small groups. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 522–523.
- Roux, L., Hu, B., Eichler, R., Stark, E., & Buzsáki, G. (2017). Sharp wave ripples during learning stabilize the hippocampal spatial map. *Nature neuroscience*, 20(6), 845–853.
- Ruusuvirta, T., Korhonen, T., Penttonen, M., Arikoski, J., & Kivirikko, K. (1995). Behavioral and hippocampal evoked responses in an auditory oddball situation when an unconditioned stimulus is paired with deviant tones in the cat: Experiment ii. *International journal of psychophysiology*, 20(1), 41–47.

- Sanchez-Fibla, M., Bernardet, U., Wasserman, E., Pelc, T., Mintz, M., Jackson, J. C., Lansink, C., Pennartz, C., & Verschure, P. F. (2010). Allostatic control for robot behavior regulation: A comparative rodent-robot study. *Advances in Complex Systems*, 13(03), 377–403.
- Sandoval, J., Su, H., Vieyres, P., Poisson, G., Ferrigno, G., & De Momi, E. (2018). Collaborative framework for robot-assisted minimally invasive surgery using a 7-dof anthropomorphic robot. *Robotics and Autonomous Systems*, 106, 95–106.
- Santos-Pata, D., Amil, A. F., Raikov, I. G., Rennó-Costa, C., Mura, A., Soltesz, I., & Verschure, P. F. (2021a). Entorhinal mismatch: A model of self-supervised learning in the hippocampus. *Isience*, 24(4).
- Santos-Pata, D., Amil, A. F., Raikov, I. G., Rennó-Costa, C., Mura, A., Soltesz, I., & Verschure, P. F. (2021b). Epistemic autonomy: Self-supervised learning in the mammalian hippocampus. *Trends in cognitive sciences*, 25(7), 582–595.
- Schrodinger, E. (1944). *What is life?: With mind and matter and autobiographical sketches*. Cambridge university press.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1), 11.
- Semba, K., & Fibiger, H. C. (1992). Afferent connections of the laterodorsal and the pedunculopontine tegmental nuclei in the rat: A retro-and antero-grade transport and immunohistochemical study. *Journal of Comparative Neurology*, 323(3), 387–410.
- Senft, E., Lemaignan, S., Baxter, P. E., Bartlett, M., & Belpaeme, T. (2019). Teaching robots social autonomy from in situ human guidance. *Science Robotics*, 4(35).
- Shein-Idelson, M., Ondracek, J. M., Liaw, H.-P., Reiter, S., & Laurent, G. (2016). Slow waves, sharp waves, ripples, and rem in sleeping dragons. *Science*, 352(6285), 590–595.
- Sherrington, C. S. (1906). *The integrative action of the nervous system* (Vol. 35). Yale University Press.
- Shum, M., Kleiman-Weiner, M., Littman, M. L., & Tenenbaum, J. B. (2019). Theory of minds: Understanding behavior in groups through inverse planning. *arXiv preprint arXiv:1901.06085*.
- Skaggs, W., McNaughton, B., & Gothard, K. (1992). An information-theoretic approach to deciphering the hippocampal code. *Advances in neural information processing systems*, 5.

- Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6(2), 149–172.
- Solé, R., Moses, M., & Forrest, S. (2019). Liquid brains, solid brains [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1774), 20190040. <https://doi.org/10.1098/rstb.2019.0040>
- Stellar, E. (1954). The physiology of motivation. *Psychological review*, 61(1), 5.
- Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & behavior*, 106(1), 5–15.
- Sterling, P. (2020). *What is health?: Allostasis and the evolution of human design*. MIT Press.
- Sterling, P., & Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. In S. Fisher & J. Reason (Eds.), *Handbook of life stress, cognition and health* (pp. 629–649). John Wiley & Sons.
- Stock, J. B., & Baker, M. (2009). Chemotaxis. In *Encyclopedia of microbiology, third edition* (pp. 71–78). Elsevier.
- Stopper, C. M., & Floresco, S. B. (2014). What's better for me? fundamental role for lateral habenula in promoting subjective decision biases. *Nature neuroscience*, 17(1), 33–35.
- Strecker, R., Nalwalk, J., Dauphin, L., Thakkar, M., Chen, Y., Ramesh, V., Hough, L., & McCarley, R. (2002). Extracellular histamine levels in the feline preoptic/anterior hypothalamic area during natural sleep–wakefulness and prolonged wakefulness: An in vivo microdialysis study. *Neuroscience*, 113(3), 663–670.
- Sturluson, S., & Brodeur, A. G. (2006). *The prose edda: Tales from norse mythology*. Courier Corporation.
- Sutton, R. S., Barto, A. G., et al. (1998). *Reinforcement learning: An introduction* (Vol. 1). MIT press Cambridge.
- Suzuki, H., & Wood, R. J. (2020). Origami-inspired miniature manipulator for tele-operated microsurgery. *Nature machine intelligence*, 2(8), 437–446.
- Szymusiak, R. (2018). Body temperature and sleep. *Handbook of clinical neurology*, 156, 341–351.
- Takahashi, Y., Zhang, Z., Kahnt, T., & Schoenbaum, G. (2024). Dopaminergic responses to identity prediction errors depend differently on the orbitofrontal cortex and hippocampus. *bioRxiv*, 2024–12.
- Thomason, J., Padmakumar, A., Sinapov, J., Walker, N., Jiang, Y., Yedidsion, H., Hart, J., Stone, P., & Mooney, R. J. (2019). Improving grounded natural language understanding through human-robot dialog. *2019 International Conference on Robotics and Automation (ICRA)*, 6934–6941.

- Thornton, E., & Evans, J. (1984). The effects of lesions of the habenula nucleus on lever press behaviour during a tandem operant schedule with contrasting response requirements. *Behavioural brain research*, 12(3), 327–334.
- Tian, J., & Uchida, N. (2015). Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. *Neuron*, 87(6), 1304–1316.
- Tingley, D., McClain, K., Kaya, E., Carpenter, J., & Buzsáki, G. (2021). A metabolic function of the hippocampal sharp wave-ripple. *Nature*, 597(7874), 82–86.
- Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science*, 307(5715), 1642–1645.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, 55(4), 189.
- Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard university press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, 28(5), 675–691.
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current anthropology*, 53(6), 673–692.
- Tranzatto, M., Dharmadhikari, M., Bernreiter, L., Camurri, M., Khattak, S., Mascarich, F., Pfreundschuh, P., Wisth, D., Zimmermann, S., Kulkarni, M., et al. (2024). Team cerberus wins the darpa subterranean challenge: Technical overview and lessons learned. *Field Robotics*, 4, 349–312.
- Tschantz, A., Barca, L., Maisto, D., Buckley, C. L., Seth, A. K., & Pezzulo, G. (2022). Simulating homeostatic, allostatic and goal-directed forms of interoceptive control using active inference. *Biological Psychology*, 169, 108266.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological review*, 108(3), 550.
- Van Der Meer, M. A., & Redish, A. D. (2009). Covert expectation-of-reward in rat ventral striatum at decision points. *Frontiers in integrative neuroscience*, 3, 458.
- Van Der Meer, M. A., & Redish, A. D. (2011). Theta phase precession in rat ventral striatum links place and reward information. *Journal of neuroscience*, 31(8), 2843–2854.
- Van Wynsberghe, A. (2020). Designing robots for care: Care centered value-sensitive design. In *Machine ethics and robot ethics* (pp. 185–211). Routledge.

- Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, 5(4), 187–196.
- Vargas, R., Porsteinsson, H., & Karlsson, K. (2012). Spontaneous neural activity of the anterodorsal lobe and entopeduncular nucleus in adult zebrafish: A putative homologue of hippocampal sharp waves. *Behavioural brain research*, 229(1), 10–20.
- Vernon, D. (2022). Cognitive architectures. *Cognitive Robotics*. MIT Press, Cambridge.
- Verschure, P., Halloy, J., & Prescott, T. (2020). Sapiens 5.0: A manifesto for the development of human-ai collaboration for good. In *Futures studies* (pp. 83–110). The Emirates Center for Strategic Studies; Research.
- Verschure, P. F. (2012). Distributed adaptive control: A theory of the mind, brain, body nexus. *Biologically Inspired Cognitive Architectures*, 1, 55–72.
- Verschure, P. F. (2016). Synthetic consciousness: The distributed adaptive control perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1701), 20150448.
- Verschure, P. F., Kröse, B. J., & Pfeifer, R. (1992). Distributed adaptive control: The self-organization of structured behavior. *Robotics and Autonomous Systems*, 9(3), 181–196.
- Verschure, P. F., Pennartz, C. M., & Pezzulo, G. (2014). The why, what, where, when and how of goal-directed choice: Neuronal and computational principles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), 20130483.
- Verschure, P. F., Voegtlin, T., & Douglas, R. J. (2003). Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, 425(6958), 620–624.
- Vinograd, A., Nair, A., Kim, J. H., Linderman, S. W., & Anderson, D. J. (2024). Causal evidence of a line attractor encoding an affective state. *Nature*, 634(8035), 910–918.
- Vouloutsi, V., Chesson, A., Blancas, M., Guerrero, O., & Verschure, P. F. (2020). The use of social sensorimotor contingencies in humanoid robots. *Biomimetic and Biohybrid Systems: 9th International Conference, Living Machines 2020, Freiburg, Germany, July 28–30, 2020, Proceedings 9*, 378–389.
- Vouloutsi, V., Lallée, S., & Verschure, P. F. (2013). Modulating behaviors using allostatic control. *Biomimetic and Biohybrid Systems: Second International Conference, Living Machines 2013, London, UK, July 29–August 2, 2013. Proceedings 2*, 287–298.

- Vysocky, A., & Novak, P. (2016). Human - Robot collaboration in industry. *MM Science Journal*, 2016, 903–906. https://doi.org/10.17973/MMSJ.2016_06_201611
- Walter, W. G. (1950). An imitation of life. *Scientific american*, 182(5), 42–45.
- Wang, Q., Jiao, W., Wang, P., & Zhang, Y. (2020). Digital twin for human-robot interactive welding and welder behavior analysis. *IEEE/CAA Journal of Automatica Sinica*, 8(2), 334–343.
- Wang, X., Shen, H., Yu, H., Guo, J., & Wei, X. (2022). Hand and arm gesture-based human-robot interaction: A review. *Proceedings of the 6th International Conference on Algorithms, Computing and Systems*, 1–7.
- Webots. (n.d.). [Http://www.cyberbotics.com](http://www.cyberbotics.com) (C. Ltd., Ed.) [Open-source Mobile Robot Simulation Software]. <http://www.cyberbotics.com>
- Weiss, A., Wortmeier, A.-K., & Kubicek, B. (2021). Cobots in Industry 4.0: A Roadmap for Future Practice Studies on Human–Robot Collaboration [Conference Name: IEEE Transactions on Human-Machine Systems]. *IEEE Transactions on Human-Machine Systems*, 51(4), 335–345. <https://doi.org/10.1109/THMS.2021.3092684>
- Wiener, N. (1948). *Cybernetics* john wiley & sons. Inc. New York.
- Wikenheiser, A. M., & Schoenbaum, G. (2016). Over the river, through the woods: Cognitive maps in the hippocampus and orbitofrontal cortex. *Nature Reviews Neuroscience*, 17(8), 513–523.
- Wilhelm, J., Petzoldt, C., Beinke, T., & Freitag, M. (2021). Review of digital twin-based interaction in smart manufacturing: Enabling cyber-physical systems for human-machine interaction. *International Journal of Computer Integrated Manufacturing*, 34(10), 1031–1048.
- Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1), 1–24.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, 9, 625–636.
- Wilson, S. P., & Prescott, T. J. (2022). Scaffolding layered control architectures through constraint closure: Insights into brain evolution and development. *Philosophical Transactions of the Royal Society B*, 377(1844), 20200519.
- Xu, W. (2019). Toward human-centered ai: A perspective from human-computer interaction. *interactions*, 26(4), 42–46.
- Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., Merrifield, R., et al. (2018). The grand challenges of science robotics. *Science robotics*, 3(14), eaar7650.
- Yepes, J. C., Yepes, J. J., Martinez, J. R., & Pérez, V. Z. (2013). Implementation of an android based teleoperation application for controlling a kuka-kr6

- robot by using sensor fusion. *2013 Pan American Health Care Exchanges (PAHCE)*, 1–5.
- Yoshida, N., Daikoku, T., Nagai, Y., & Kuniyoshi, Y. (2024). Emergence of integrated behaviors through direct optimization for homeostasis. *Neural Networks*, 177, 106379.
- Young, J. J., & Shapiro, M. L. (2011). Dynamic coding of goal-directed paths by orbital prefrontal cortex. *Journal of Neuroscience*, 31(16), 5989–6000.
- Yuan, Y., Wu, W., Chen, M., Cai, F., Fan, C., Shen, W., Sun, W., & Hu, J. (2019). Reward inhibits paraventricular crh neurons to relieve stress. *Current Biology*, 29(7), 1243–1251.
- Zimmerman, C. A., Leib, D. E., & Knight, Z. A. (2017). Neural circuits underlying thirst and fluid homeostasis. *Nature Reviews Neuroscience*, 18(8), 459–469.

APPENDIX C – Curriculum Vitae

Oscar Guerrero Rosado (b. 1992, Ronda, Spain) is a Ph.D. candidate at the Donders Institute for Brain, Cognition and Behaviour at Radboud University, where he investigates the neural mechanisms underlying motivated behavior, and how these can guide the design of intrinsically motivated artificial agents. His work focused on biologically constrained models of allostatic regulation, hippocampal representation, and reward-driven learning, which are implemented and tested in both simulated and physical robotic platforms.

Oscar began his research career during his Psychology degree at the University of Granada, where he explored how abstract concepts, such as time, are interpreted from an embodied, egocentric perspective. This interest in embodiment was later extended during his Master's in Cognitive Systems and Interactive Media at Universitat Pompeu Fabra, where he developed an interactive trampoline-based game and analyzed how participants experienced game difficulty as a function of the behavioral repertory available to actuate in the video game.

After completing his Master's, Oscar joined the Synthetic Perceptive, Emotive and Cognitive Systems (SPECS) group at the Institute for Bioengineering of Catalonia (IBEC), first as a research assistant and later as a Ph.D. candidate in collaboration with Universitat Pompeu Fabra. Within this environment, he contributed to several European projects, including HR-Recycler, where he worked on human-robot collaboration in recycling plants, and CAVAA, which extended motivational models toward a broader theory of synthetic awareness.

Oscar has been an active member of the international research community presenting his work at major conferences such as IROS (*International Conference on Intelligent Robots and Systems*) and ALIFE (*Artificial Life*). He has also contributed to advanced training programs and summer schools, including *Living Machines*, *BCBT/DCBT*, *Neuromatch Academy*, and the *MRS Summer School on Multi-Robot Systems*, which has broadened his expertise in computational neuroscience, machine learning, and robotics.

His academic trajectory combines psychology, neuroscience, artificial intelligence and robotics, with the goal of understanding how purposive and adaptive behavior emerges from the interplay between biological needs and environmental opportunities. Inspired by this interdisciplinary perspective, his research advocates for a new design paradigm in artificial intelligence and robotics, where autonomy arises not from externally defined objectives but from the intrinsic regulation of needs.

Publications included in this thesis

- Guerrero-Rosado, O.**, Amil, A. F., Freire, I. T., & Verschure, P. F. (2022). *Drive competition underlies effective allostatic orchestration*. *Frontiers in Robotics and AI*, 9, 1052998.
- Guerrero-Rosado, O.**, Amil, A. F., Freire, I. T., Vinck, M., & Verschure, P. F. (2025). *Biomimetic self-regulation in intrinsically motivated robots*. In *Artificial Life Conference Proceedings 37* (Vol. 2025, No. 1, p. 49). One Rogers Street, Cambridge, MA 02142-1209, USA journals-info@mit.edu: MIT Press.
- Guerrero-Rosado, O.**, Amil, A. F., Freire, I. T., Vinck, M., & Verschure, P. F. (2025). *Motivational Cognitive Maps Allow Robot Biomimetic Autonomy*. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 14402-14408). IEEE.
- Freire, I. T. [†], **Guerrero-Rosado, O.** [†], Amil, A. F., & Verschure, P. F. (2024). *Socially adaptive cognitive architecture for human-robot collaboration in industrial settings*. *Frontiers in Robotics and AI*, 11, 1248646.

Other publications

- Guerrero-Rosado, O.**, & Verschure, P. F. (2020). *Distributed adaptive control: An ideal cognitive architecture candidate for managing a robotic recycling plant*. In *Conference on biomimetic and biohybrid systems* (pp. 153-164).
- Vouloutsis, V., Chesson, A., Blancas, M., **Guerrero-Rosado, O.**, & Verschure, P. F. (2020). *The use of social sensorimotor contingencies in humanoid robots*. In *Conference on Biomimetic and Biohybrid Systems* (pp. 378-389).
- Guerrero-Rosado, O.**, & Verschure, P. (2021). *Robot regulatory behaviour based on fundamental homeostatic and allostatic principles*. *Procedia Computer Science*, 190, 292-300.
- Ngo, T., **Guerrero-Rosado, O.**, Freire, I. T., & Verschure, P. F. (2022). *Homeostatic and allostatic principles for behavioral regulation in desert reptiles: a robotic evaluation*. In *Conference on Biomimetic and Biohybrid Systems* (pp. 332-344).
- Kahali, S., Ngo, T., Mandahar, T. S., Martínez-Bueno, E., **Guerrero-Rosado, O.**, López-Carral, H., & Verschure, P. F. (2023). *Distributed adaptive control for virtual cyborgs: a case study for personalized rehabilitation*. In *Conference on Biomimetic and Biohybrid Systems* (pp. 16-32).

APPENDIX E – Data Management Plan

The experiments reported in this thesis are purely computational, and no animal or human subjects were involved at any stage. The data does not represent any ethical or moral issue, and its management was performed in accordance with the FAIR principles.

Rather than storing and sharing raw data, the full methodology and documentation of each chapter are publicly available at the repository links provided below, and readers can generate their own data to replicate the results.

Chapter 2. Drive Competition Underlies Effective Allostatic Orchestration.

All code required to reproduce the main results of this chapter is publicly available at:

<https://gitlab.com/OscarGuerreroRosado/allostatic-orchestration>

Chapter 3. Biomimetic Self-Regulation in Intrinsically Motivated Robots.

All code required to reproduce the main results of this chapter is publicly available at:

https://gitlab.com/OscarGuerreroRosado/alife-desert_lizard

Chapter 4. Motivational Cognitive Maps Allow Robot Biomimetic Autonomy. All code required to reproduce the main results of this chapter is publicly available at:

https://gitlab.com/OscarGuerreroRosado/motivational_hippocampal_autoencoder

Chapter 5. Socially Adaptive Cognitive Architecture for HRC in Industrial Settings. No experimental subjects were involved in this work. Personnel testing the solutions developed in this chapter were part of the HR-Recycler European project; their involvement is acknowledged at the end of the chapter.

Chapter 6. Unifying Hippocampal-VTA Roles during Goal-Directed Behavior.

This chapter reports preliminary data that has not yet been published. The corresponding repository will be made publicly available once this work is developed enough to be submitted for publication.

APPENDIX F – Donders Graduate School

For a successful research Institute, it is vital to train the next generation of scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School in 2009. The mission of the Donders Graduate School is to guide our graduates to become skilled academics who are equipped for a wide range of professions. To achieve this, we do our utmost to ensure that our PhD candidates receive support and supervision of the highest quality.

Since 2009, the Donders Graduate School has grown into a vibrant community of highly talented national and international PhD candidates, with over 500 PhD candidates enrolled. Their backgrounds cover a wide range of disciplines, from physics to psychology, medicine to psycholinguistics, and biology to artificial intelligence. Similarly, their interdisciplinary research covers genetic, molecular, and cellular processes at one end and computational, system-level neuroscience with cognitive and behavioural analysis at the other end. We ask all PhD candidates within the Donders Graduate School to publish their PhD thesis in the Donders Thesis Series. This series currently includes over 600 PhD theses from our PhD graduates and thereby provides a comprehensive overview of the diverse types of research performed at the Donders Institute. A complete overview of the Donders Thesis Series can be found on our website: <https://www.ru.nl/donders/donders-series>

The Donders Graduate School tracks the careers of our PhD graduates carefully. In general, the PhD graduates end up in high-quality positions in different sectors, for a complete overview see <https://www.ru.nl/donders/destination-our-former-phd>. A large proportion of our PhD alumni continue in academia (>50%). Most of them first work as a postdoc before growing into more senior research positions. They work at top institutes worldwide, such as University of Oxford, University of Cambridge, Stanford University, Princeton University, UCL London, MPI Leipzig, Karolinska Institute, UC Berkeley, EPFL Lausanne, and many others. In addition, a large group of PhD graduates continue in clinical positions, sometimes combining it with academic research. Clinical positions can be divided into medical doctors, for instance, in genetics, geriatrics, psychiatry, or neurology, and in psychologists, for instance as healthcare psychologist, clinical neuropsychologist, or clinical psychologist. Furthermore, there are PhD graduates who continue to work as researchers outside academia, for instance at non-profit or government organizations, or in pharmaceutical companies. There are also PhD graduates who work in education, such as teachers in high school, or as lecturers in higher education. Others continue in a

wide range of positions, such as policy advisors, project managers, consultants, data scientists, web or software developers, business owners, regulatory affairs specialists, engineers, managers, or IT architects. As such, the career paths of Donders PhD graduates span a broad range of sectors and professions, but the common factor is that they almost all have become successful professionals.

For more information on the Donders Graduate School, as well as past and upcoming defences, please visit: <http://www.ru.nl/donders/graduate-school/phd/>

Life has evolved countless strategies to resist entropy. Even the simplest organisms regulate temperature, energy, and osmolarity to preserve their fragile order. Through evolution, these regulatory mechanisms have been refined and combined, enabling life to flourish even in the most challenging environments.

This thesis takes inspiration from one of the most sophisticated control systems in nature, the mammalian brain, to translate principles of biological autonomy into artificial machines. The result is a vision of robots endowed with a sense of purpose: Agents that act not merely to execute external commands, but to sustain themselves.

Three essential features are explored, modeled, and implemented. First, the embodiment and orchestration of innate internal needs. Second, the shaping of environmental evaluation through the agent's own internal state. And third, the role of memory in preserving experiences vital for survival.

Taken together, the contributions of this thesis outline a design paradigm in which autonomy is not an afterthought but the foundational principle of control from which complex behavior naturally emerges. This perspective not only advances the science of intelligent machines but also raises pressing moral and philosophical questions about the nature of artificial purpose. Questions that demand reflection across science, philosophy, and society.