Exploring the role of neural network architecture onto decision-making processes with reservoir computing

Naomi Chaix-Eichel

Une thèse présentée pour l'obtention du diplôme de
DOCTEUR EN SCIENCES
(Dr. sc. Université de Bordeaux)

École doctorale de mathématiques et informatique
Spécialité : Informatique

Etude du rôle de l'architecture des réseaux neuronaux dans la prise de décision à l'aide de modèles de reservoir computing

Exploring the role of neural network architecture onto decision-making processes with reservoir computing

Présentée par

Naomi Chaix-Eichel

Sous la direction de

Dr. Nicolas P. Rougier, Dr. Thomas Boraud

Défendue le 22 Novembre 2024

Membres du jury :

Mme. Eleni Vasilaki	Professeur	University of Sheffield	Rapporteure
M. Peter Dominey	Directeur de Recherche	CNRS	Raporteur
M. Clément Moulin-Frier	Assistant Professeur	INRIA	Examinateur
Mme Anne Collins	Assistant Professeur	University of Berkeley	Examinatrice
M. Mehdi Khamassi	Directeur de Recherche	CNRS	Examinateur
M. Thomas Boraud	Directeur de Recherche	CNRS	Directeur de these
M. Nicolas P. ROUGIER	Directeur de Recherche	INRIA	Directeur de these

Acknowledgments

I am sincerely grateful to Nicolas P. Rougier and Thomas Boraud for their supervision over the past three years. They provided me the opportunity to immerse myself in the field of computational neuroscience, which has raised my curiosity for the future and enriched me both intellectually and personally. Their trust, support, and flexibility have been crucial throughout this journey.

I extend my deep gratitude to the members of my PhD jury for their time, effort and engagement in reviewing my thesis.

I am grateful to the people I have had the privilege to collaborate with: Snigdha Dagar, Gautham Venugopal, Aude Retailleau, Sebastien Ballesta, Ayrton Guerillon, and Yannis Bendi-Ouis. These collaborations enriched my research journey.

My appreciation extends to the Mnemosyne team for creating a pleasant and inclusive environment. Deep thanks to the IMN team for bringing joy to the lab and for the good times we've shared together, including our time in Vienna. Special thanks go to Thibault Dhellemmes and Nathan Trouvain, for their support, but also for the fun moments we've shared together. An other special mention goes to Claire Delattre for her positive presence, not just in the lab but also beyond - our several running, swimming, and cycling sessions, culminating in the triathlon we completed together, are moments I will keep in mind.

I am deeply grateful to Alexis Dubreuil for his invaluable presence in the lab since the very beginning of our journey at the IMN, which we started together. These moments and discussions we've shared together turned us into not just colleagues but friends.

I am also thankful to the OCNC team for our unforgettable experience during the month we spent in Okinawa. I am especially thankful to have bonded with the inspiring Tom George, Heloisa Chiossi, Gaston Sivori, and Hugo Musset.

Beyond the academic realm, I am thankful to my athletics team, who played an essential role throughout these three years of PhD. Special thanks to my coaches, Jean-Louis Alliot and Arnaud Lopez for their guidance and mentorship.

I extend my gratitude to my closest friends for their unwavering support. Special thanks to the women, my dear friends from Nice and Supoptique, whose strength and trust have uplifted me.

Lastly, and most importantly, I am deeply grateful to my family, especially to my incredible brothers, Kennoské and Enzo, my mother Yuri, and my dear partner, Lucas, who has been by my side, supporting me closely from the very first day of

this thesis journey.

Abstract

A striking similarity exists in the organization and structure of certain brain regions across diverse species. For instance, the brain structure of vertebrates, from fish to mammals, includes regions like the cortex, hippocampus, cerebellum and basal ganglia with remarkable similarity. The presence of these structures across a wide range of species strongly suggests that they emerged early in vertebrate evolution and have been conserved throughout evolution. The persistence of these structures raises intriguing questions about their evolutionary origins: are they unique and optimal solutions for processing information and controlling behavior, or could alternative brain architectures emerge to achieve similar functional properties? To investigate this question, this thesis explores the relationship between brain architecture and cognitive function, with a focus on decision-making processes. We propose to use variants of a recurrent neural network model (echo state network) that is structurally minimal and randomly connected. We aim to identify whether a minimal model can capture any decision-making process and if it cannot, we explore whether multiple realizable solutions emerge through structural variations. First we demonstrate that a minimal model is able to solve simple decision tasks in the context of spatial navigation. Second, we show that this minimal structure has performance limitations when handling more complex tasks, requiring additional structural constraints to achieve better results. Third, by employing a genetic algorithm to evolve network structure to more complex ones, we discover that multiple realizable solutions emerging through structural variations. Furthermore we reveal that identical architectures can exhibit a range of different behaviors, leading us to investigate additional factors contributing to these different behaviors beyond structural variations. Our analysis of the behavior of 24 monkeys living in a community reveals that social factors, such as social hierarchy, play a significant role in their behavior. This thesis takes an approach that differs from traditional neuroscience methodologies. Rather than directly constructing biologically inspired architectures, the models are designed from simple to complex structures, reproducing the process of biological evolution. By leveraging the principles of multiple realizability, this approach enables the evolution of diverse structural configurations that can achieve equivalent functional outcomes.

Résumé

Une similarité frappante existe dans l'organisation et la structure de certaines régions du cerveau chez diverses espèces. Par exemple, la structure cérébrale des vertébrés, des poissons aux mammifères, présente une similarité remarquable dans des régions telles que le cortex, l'hippocampe, le cervelet et les ganglions de la base. Cela suggère que ces régions sont apparues tôt dans l'évolution des vertébrés et ont été conservées au fil du temps. La persistance de ces structures soulève des questions fondamentales sur leurs origines évolutives : sont-elles des solutions uniques et optimales pour le traitement de l'information et le contrôle du comportement, ou d'autres architectures cérébrales pourraient-elles émerger pour offrir des propriétés fonctionnelles équivalentes ? Cette thèse étudie la relation entre la structure du cerveau et les fonctions cognitives, en se concentrant particulièrement sur le processus de prise de décision. Nous proposons d'utiliser un type de réseau de neurones récurrent appelé Echo State Network qui est structurellement minimal et dans lequel les neurones sont connectées de manière aléatoire. Nous voulons déterminer si ce modèle minimal peut capturer tout processus décisionnel et si ce n'est pas le cas, nous chercherons l'existence de structures alternatives. Premièrement, nous démontrons qu'un modèle minimal est capable de résoudre des tâches décisionnelles simples dans le contexte de la navigation spatiale. Ensuite, nous montrons que cette structure minimale a des limitations de performance lorsqu'il s'agit de tâches plus complexes, nécessitant plus de structures pour retrouvant de bonnes performances. Troisièmement, nous utilisons un algorithme génétique faisant évoluer la structure du réseau vers des configurations plus complexes, ce qui nous conduit à découvrir plusieurs solutions réalisables émergeantes de variations structurelles. De plus, nos résultats révèlent que des architectures identiques peuvent manifester une gamme de comportements différents, nous incitant à explorer les facteurs supplémentaires pouvant contribuer à ces différences comportementales, au-delà des variations structurelles. Notre analyse du comportement de 24 singes vivant en communauté révèle que des facteurs sociaux, tels que la hiérarchie sociale, jouent un rôle significatif dans l'influence du comportement. Cette thèse adopte une approche qui diffère des méthodologies traditionnelles en neurosciences. Plutôt que de construire directement des architectures biologique, les modèles sont construits en faisant évoluer leur structures de simples à complexes, reproduisant ainsi le processus de l'évolution biologique. En s'appuyant sur les principes de réalisabilité multiple, cette approche permet l'évolution de configurations structurelles diverses capables de parvenir à des résultats fonctionnels équivalents.

Contents

Introduction

Over the past decades, understanding the relationship between brain structures and functions has been a central focus of neuroscience research. Brain connectivity is heavily studied in order to understand how different brain regions and neurons are connected and interact with each other. Fundamental questions are addressed: how does the complex architecture of the brain give rise to cognitive abilities ? And what are the consequences of alterations in brain structure on cognitive functions ? To this end, ongoing research aims to unravel the brain's organization and its functions, leading researchers to view the brain as a network [Sporns, 2016], comprising multiple units and modules that are intricately connected. The relationships between structure, connectivity, and computation [Ostojic and Fusi, 2024] is critical for understanding how the brain processes and transmit information [Sporns et al., 2005].

Mapping the structure of the brain network, referred as the connectome [Sporns et al., 2005], is crucial to understand the functional implication of the structure. So far, only the C. elegans connectome and the Drosophila melanogaster (fruit fly) connectome is fully known [Cook et al., 2019, Schlegel et al., 2024]. Nevertheless, advances in neuroimaging, neuro-electrophisology and computational modeling enabled increased research application to connectomes. For instance, image techniques like Diffusion Tensor Imaging (DTI) [Hagmann et al., 2008] and functional MRI (fMRI) [Smith et al., 2013] are employed to measure brain activity during a task or at rest, representing powerful tools to study functional connectivity. Additionally, a large scale project called the Human Connectome Project (HCP) [Van Essen et al., 2012], which aims at mapping the entire human connectome, is also ongoing in the direction of unravelling brain connectivity.

Research has shown that the brain connectivity exhibits a combination of random and non random properties [Esposito et al., 2014]. This mix is reminiscent of small-world networks [Watts and Strogatz, 1998], which have been found to be a property of brain connectivity [Bassett and Bullmore, 2006, Liao et al., 2017]. Small-world networks comprise random connections that facilitate shortcuts between distant nodes, reducing average path length. Additionally, they exhibit a mix of regularity and local clustering, where neurons tend to form clusters and are more likely to connect with neighboring nodes than randomly selected ones. Both random and non-random properties provide a balance between local clustering and global connectivity, rendering them efficient for information transmission. The small-world property of the brain have been extensively studied and empirically confirmed in

the case of Caenorhabditis Elegans [Watts and Strogatz, 1998], nevertheless still discussed and reconsidered [Hilgetag and Goulas, 2016, Muller et al., 2014]. Furthermore, research has also revealed the presence of modular properties in the brain [Meunier et al., 2010, Sporns and Betzel, 2016], characterized by a type of neural organization where the brain is divided into distinct modules that specialize for specific functions. Each module can be considered as a subnetwork comprising interconnected neurons that collaborate to perform a specific task, and these modules communicate with each other through dedicated connections. Additionally, several other important connectivity properties are observed in the brain, including scale-free network [Eguiluz et al., 2005] and hierarchical organization [Zhou et al., 2006].

Knowing the human connectome is not only crucial to understand the implications of structure on healthy brains but also dysfunctional ones. Indeed, small-world properties may be essential for maintaining healthy brain function [Liao et al., 2010] since alterations of small-world features are observed in patient groups with Alzheimer's disease [Stam et al., 2007], autism [Barttfeld et al., 2012] and schizophrenia [Liu et al., 2008, Fornito et al., 2011, Micheloyannis et al., 2006]. Moreover, it has been shown that structural changes were occurring during aging [Fjell and Walhovd, 2010], and that epileptic patients showed an altered modular organization [Chavez et al., 2010].

Meanwhile, the origins of these specific connectivity properties of the brain are still questioned. What drives the brain to adopt this particular organization rather than another? Sporns [2011] suggest that non-random properties of the brain may arise from spatial embedding and wiring economy considerations. Random connectivity has a high cost as it does not consider physical constraints, whereas distance-dependent and cluster organization allow a wiring economy, minimizing unnecessary neural connections. Another suggestion is that these specific structures are essential for the emergence of certain cognitive functions. This idea is supported by the striking similarities observed in the organization of brain regions across diverse species [Ardesch et al., 2019, Van den Heuvel et al., 2016, Woych et al., 2022, Kozol et al., 2023]. Such consistency over millions of years of evolution suggests that these structures may be crucial for specific cognitive functions.

In this study, we investigate the impact of brain structure on a particular cognitive function: decision-making. The latter refers to the cognitive process by which an individual has to select one option among others based on several sensory information, preferences, goals and past experiences. The Basal-Ganglia (BG) have been shown to play a crucial role in this cognitive process [Boraud, 2020]. It corresponds to a group of subcortical structures in the vertebrate that has remarkably preserved its structure throughout 500 millions years of evolution [Grillner et al., 2013]. This raises questions about the origin of the conservation. Is this specific structure essential for action selection and decision-making across vertebrate species? This question serves as the primary motivation for our work. By uncovering potential alternatives

to the traditional focus on BG anatomy, we aim to explore the broader impact of neural structure on decision-making networks by investigating whether other structures can also influence this cognitive process.

Our approach employs a strategy that diverges from conventional modeling methods: we analyze a system without any preconceived structure. Instead of constructing models inspired by biological decision-making networks, we don't impose assumptions about necessary structural components. This enables us to identify the minimal requirements for effective decision-making and to discern which features emerge naturally from the behavior and learning process, rather than being built into the model a priori. Consequently, we can explore whether complex cognitive functions can arise from simpler architectures.

We first explore whether an unstructured network can effectively learn simple decision-making. We secondly investigate whether the same network can tackle more complex decision tasks, and observe that it cannot achieve good performance without additional structure. This brings us to our third question: if a structured network is needed for complex decision-making, could there be multiple structures that achieve this? This idea relates to multiple realizability, which suggests that mental states can be supported by different neural mechanisms rather than just one. Multiple realizability was first introduced by Putnam [1967], who used the example of pain: both humans and animals can experience pain, but the underlying neural processes might differ. In the context of multiple realizability, understanding how multiple functions can be realized by different structures is studied under the term of *degeneracy* [Edelman and Gally, 2001]. Degeneracy refers to the ability of structurally different elements to perform the same function. This property is common in biological systems and has become a topic of significant interest. A notable example can be found in the comparison between the decision-making systems of octopuses and vertebrates. Thanks to the highly distributed system of the octopus, its arms can make decisions independently of its brain due to the presence of neurons and ganglia along the arms [Hochner, 2012]. In contrast, the decision-making system in vertebrates is centralized, with the basal ganglia playing a crucial role. This illustrates how two very different systems can both achieve highly efficient decision-making. In this thesis, we investigate degeneracy in the context of economic decision-making, focusing on how various brain structures contribute to varying behaviors. Economic decision-making involves assessing subjective values of options, selecting one among them, evaluating the outcome, and learning from it to improve future decisions. We aim to determine the extent to which different brain structures can perform these steps. Specifically, we'll look at subjective choices in risky situations to study risk-taking behavior- whether there's a tendency toward risk aversion or risk-seeking. Risk attitudes are fundamental to study behavior across species, potentially driving evolutionary processes. With this in mind, and to avoid making strong assumptions about the architecture, we're designing the network structures using evolutionary principles. We begin with an initial, unstructured, and simple network, allowing it

to evolve according to evolutionary rules like reproduction, recombination, mutation and selection. This approach mimics natural evolution, where successful structures emerge without preconceived knowledge about its optimal configuration. *We will demonstrate that multiple structures can produce the same behavior and also that similar structures can yield different behaviors.*

The modeling part of this work is carried out using mathematical modeling to simulate and train a neural network on decision-making tasks. We use Recurrent Neural Networks (RNNs), a type of artificial neural network widely used in computational neuroscience [Sussillo, 2014] for solving cognitive tasks [Yang and Molano-Mazón, 2021, Barak, 2017]. RNNs are known to have strong similarities to the biological brain, with the presence of non linear units, recurrent neural connections, and the ability to exhibit a similar complex internal dynamics [Van Vreeswijk and Sompolinsky, 1996]. They are often used to model brain regions with known recurrent connections such as specific parts of the hippocampus (CA3, CA1, and dentrate gyrus), the cerebellum, and the prefrontal cortex. Despite these similarities, the training methods of these biological RNN are still under investigation. Training RNNs remains a challenge in the AI domain, primarily when using backpropagation, a method known for its susceptibility to issues such as the vanishing and exploding gradient problem [Pascanu, 2013]. This method is further questioned since it has been subject to debate regarding its biological plausibility [Lillicrap et al., 2020]. Consequently, researchers have explored alternative approaches such as using other training rules in a more local way, for example by using Spike-Timing-Dependent Plasticity and reward-based learning, but these methods display some limitations in terms of stability of learning and capturing the complex dynamics. One alternative we explore in this thesis is reservoir computing [Lukoševičius and Jaeger, 2009], an approach inspired by the brain's information processing capabilities. This method leverages the inherent dynamics of neural-like systems without the need for training the recurrent components. Unlike traditional RNNs in which both the recurrent connections and the output weights are trained, reservoir computing decouples the dynamics within the reservoir from the learning process of the output layer. This separation alleviates the need for training the recurrent connections and avoids the backpropagation issue. The reservoir part is composed of a large recurrent neural network with fixed connections, while the output layer maps the reservoir internal activity to the output and is the only trainable component. Reservoir computing offers high computational efficiency, because only the output layer requires training, while maintaining the benefits of RNNs, including the ability to process temporal data through recurrent connections in the reservoir, which allows for the storage and retrieval of past information. We will use a specific type of reservoir computing called Echo State Network (ESN) [Jaeger, 2007], a RNN model with randomly interconnected neurons, and its output layer trained with linear regression. Employing ESN enables us to construct models with randomly configured connectivity, mirroring certain random properties found in the brain. As ESN is randomly initialized, this model can be regarded as minimal, lacking strong prior assumptions about its

architecture. Consequently, it corresponds to our definition of a minimal models without any architectural constraints. More information about the technical properties of the ESN are explained in Boxes 0 .1 and 0 .2.

Our investigation is structured around four chapters. The first three chapters utilize this minimal ESN model, while the final chapter focuses on analyzing a large dataset of behavioral data on monkeys.

The first chapter examines whether simple decision-making can be achieved using the minimal ESN model. We chose to study decision-making in the context of spatial navigation, a process that heavily involves the hippocampus. This choice enables us to support our analysis with a wealth of existing experimental and theoretical studies on neural dynamics. We simulated an agent navigating an 8-shaped track controlled by a random ESN network. The model successfully solved an alternating task despite bearing no structural similarity to the hippocampal formation. Furthermore, we identified neural properties typically associated with the hippocampus and known to play crucial roles in spatial navigation [Moser et al., 2015, 2008]. These findings suggest that the unstructured model can effectively solve a simple navigation task, and that the observed cell activity emerges from the agent's behavior rather than from specific structural or anatomical features.

In the second chapter, we investigate whether such ESN minimal model can capture any other decision-making process. Since ESNs have been demonstrating strong performances in benchmark tasks [Lukoševičius, 2012, Jaeger, 2007, Verstraeten et al., 2006], it is legitimate to wonder whether their capabilities are sufficient to solve any task. We selected a task type where random time constraints add difficulty to achieving high performance and analyzed how structural modifications could enhance performance as task complexity increases. Our findings show the benefits of having a more structured architecture and the advantages of combining multiple ESNs in Deep ESN models [Gallicchio and Micheli, 2017]. Specifically, we demonstrate that the best-performing structures include both fast and slow pathways, mirroring a functional property of the basal ganglia, which is highly involved in decision-making. Overall, we demonstrate that while a minimal ESN model can handle simpler tasks, additional structural complexity is necessary to solve more complex tasks effectively.

In the third chapter, we investigate whether different structures can solve the same decision-making process. To achieve this, we adapt the minimal model to risky decision-making. We train the model to perform a task previously studied in monkeys and use genetic algorithms to evolve the model's architecture. We subsequently analyze the converging structures that emerge from this evolutionary process. Our results demonstrate that multiple distinct structures can arise, each achieving similar task performance, highlighting the concept of multiple realizability and degeneracy. Additionally, we demonstrate that, at the population level, these structures show

consistent trends in attitudes toward risk. However, at the individual level, networks within the same structural family can exhibit diverse risk-taking behaviors. This underscores that while structure does influence risk attitudes, leading to similar population-level risk-attitude, there remains some inter-individual variability.

In the final chapter, we focus on this inter-individual variability by questioning why similar structures can lead to diverse behaviors. To investigate this, we shift from modeling to analyzing real behavioral data from a population of 24 monkeys performing the same economic task. We ask what makes one monkey more risk-averse than another, even within the same species. Using the powerful Prospect Theory (PT) framework [Kahneman and Tversky, 1979], which effectively assesses attitudes toward risk, we quantify when monkeys are risk-averse, risk-seeking, or loss-averse, concepts observed in humans and some animals. Our analysis confirms that, overall, monkeys tend to be risk-averse in gain domains, exhibit loss aversion, and distort probabilities. However, we also find significant inter-individual differences. We show that social factors, such as hierarchical status, appear to contribute to this variability and influence each monkey's risk-taking behavior: dominant monkeys tend to be less risk-averse than their subordinates.

We conclude this manuscript by summarizing our key findings and exploring the perspectives for future research.

Box - Modeling the minimal model with the ESN paradigm - 1

An Echo State Network (ESN) is a type of recurrent neural network that belongs to the specific category of reservoir computing [Jaeger, 2007]. It is composed of three parts: the input layer, the reservoir part, and the read-out layer. The reservoir part, depicted in the center part of Figure 1.2, is composed of a pool of neurons randomly connected to themselves and to the input.

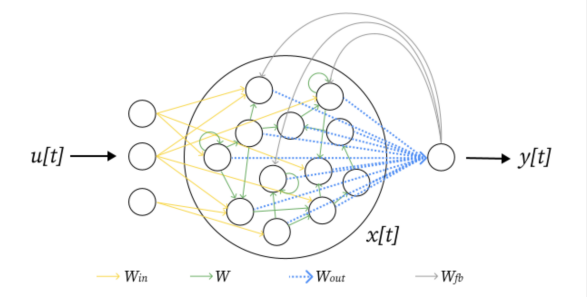

Figure 1: ESN architecture [a]
. $u(t)$, $x(t)$, $y(t)$ represent the input, reservoir state and output at time t respectively. W and W_{in} are randomly initialized then frozen, whereas W_{out} is plastic.

The neurons have the following dynamics:

$$\frac{1}{\alpha}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \tanh(W.\mathbf{x} + W_{in}.\mathbf{u} + W_{fb}.\mathbf{y}) \tag{0.1}$$

$$\mathbf{y} = W_{out}.\mathbf{x} \tag{0.2}$$

where $\mathbf{x}$, $\mathbf{u}$ and $\mathbf{y}$ represent the reservoir states, input, and output. W, W_{in}, and W_{out},W_{fb} are weight matrices, while $tanh$ refers to the hyperbolic tangent function, which we have chosen as the activation function for this thesis. However, other activation functions such as sigmoid, softmax, and softplus could also be used. α refers to the leak rate, a crucial parameter of the ESN that plays a role in controlling the memory and timescale of the network's dynamics: a small leak rate indicates a bigger memory and a slower dynamics, whereas a big leak rates lead to a smaller memory but a higher speed of update dynamics [Lukoševičius, 2012]. The reservoir computing approach has been shown to exhibit several biologically plausible characteristics [Lukoševičius and Jaeger, 2009].

[a] Figure from ReservoirPy

Box - Modeling the minimal model with the ESN paradigm - 2

These include its sparse [Manneschi et al., 2021b] and random connectivity, inspired by biological neural networks, and its recurrent connections, similar to those found in the cortex [Enel et al., 2016, Hinaut et al., 2015]. Furthermore, ESNs are energy-efficient, akin to the brain's ability to perform complex computations with minimal energy. These similarities makes ESN a valuable model for understanding certain aspects of the brain.
The input and reservoir connections, depicted in yellow and green respectively in Figure 1.2, are not trained. These connections are randomly initialized after some hyperparameter tuning. Only the output layer, referred to as the readout neurons (shown in red in Figure 1.2), is trained. This approach allows the use of very simple training rules and avoids the challenges associated with training recurrent neural networks (RNNs) using backpropagation, making the training of the ESN computationally inexpensive compared to other classical RNNs. Additionally, training ESNs requires significantly less data than training other RNNs [Hinaut and Trouvain, 2021, Variengien and Hinaut, 2020]. The most common method for training the readout layer is supervised learning using ridge regression on the training data. In this thesis, we will use these supervised method, along with more biologically plausible training methods such as reinforcement learning. All models are implemented using the Python library ReservoirPy [Trouvain et al., 2020b].

1 Chapter 1: An Unstructured Network can solve non-trivial decision task

1.1 Introduction

In this chapter, we study decision-making in the context of spatial navigation, a process that heavily involves the hippocampus[1]. The hippocampal formation has been heavily studied in the past few decades [Maguire et al., 1996, Buzsáki and Moser, 2013, Hartley et al., 2014, Moser et al., 2017] and researchers have since then established a huge repertoire of cells displaying very specific properties. In his book "The Brain from Inside out" [György Buzsáki, 2019], György Buzsáki named a few of them in a footnote (page 356): *place cells, time cells, grid cells, head direction cells, boundary vector cells, border cells, object cells, object vector cells, distance cells, reward cells, concept cells, view cells, speed cells, memory cells, goal cells, goal direction cells, splitter cells, prospective cells, retrospective cells, island cells, ocean cells, band cells, object cells, pitch cells, trial order cells, etc.* Each and every cells have been characterized in terms of correlation between their activity and some combination of high level property involving space, time and internal state. This is the case for place (cell fires when animal is in a specific place), head direction (cell fires preferentially when head is pointing toward a specific direction), time cells (cell fires at successive moments), etc. Facing such huge repertoire, one may legitimately wonder how the brain orchestrates all these information and makes use of it to ensure survival. Part of the answer is given by György Buzsáki himself in the same footnote where he further explained that *physiological attributes of neurons in the hippocampal-enthorinal system ... might be explained by the apparent distinctiveness of a few sets of continuous distributions.* In other words, even though we can and we do observe these cells *in vivo*, they might be a simple epiphenomenon: their activity might be correlated to some unknown latent variables.

This is actually hardly different from the hypothesis provided by Raju et al. [2024] where authors suggest that spatial representations are not explicitly encoded but emerge as a byproduct of sequence learning. Their model is based on a variant of Hidden Markov Model (HMM) theory and employs a clone-structured causal graph (CSCG) [George et al., 2021] to differentiate between various sequences of input. In this framework, multiple "clones" represent the same observation across

[1]This chapter is an extract of [Chaix-Echel et al., 2024]

different contexts, enabling the model to handle ambiguous sensory input sequentially. Consequently, the representation of space arises implicitly through the process of sequence learning, rather than being an explicit function of the hippocampus. Other research have similarly challenged the traditional space-centric view of the hippocampus. Using the concept of successor representation [Dayan, 1993], Stachenfeld et al. [2017] suggest that hippocampal cells encode a predictive map rather than a purely euclidean spatial map. Sanders et al. [2020] propose that the hippocampus builds more abstract representations of environmental structure based on the animal's subjective beliefs rather than directly representing objective spatial properties of the environment. Whittington et al. [2020] introduced the Tolman-Eichenbaum machine (TEM), a model that unifies spatial and non spatial functions, arguing that the emergence of the hippocampal cells emerge as a result of the system learning to represent abstract relationships and can represent both spatial and non-spatial information.

In this study, we align with current approaches and propose to reconsider the role of hippocampal cells. More precisely, we decided to focus our study on splitter cells about which Duvelle et al. [2023] published recently an extended review of the literature. Place cells are a type of hippocampal neurons that fire when an animal occupies a specific location in the environment, helping in building a cognitive representation of a specific location in space [O'Keefe and Dostrovsky, 1971]. Within the population of place cells, there exists a subset known as "splitter cells" [Wood et al., 2000], that not only encode the animal's current location similar to conventional place cells, but also encode information about the past or the future trajectory. In other words, splitter cells demonstrate different firing patterns depending on the animal's origin or intended destination, even if the animal is in the same physical location.

We adopt a novel perspective based on György Buzsáki's interpretation of the brain as a large reservoir of neurons with pre-existing patterns of activity [György Buzsáki, 2019]. According to this view, when a new experience occurs, the brain assigns it to a pre-existing pattern from a vast reservoir of internally generated patterns, thereby assigning behavioral meaning to the neuronal sequence. This mechanism is considered more biologically plausible than creating entirely new patterns for each experience. Building on this concept, we propose to study a reservoir computing model composed of a pool of recurrently interconnected neurons, representing a vast reservoir capable of generating an extensive repertoire of neuronal patterns. The model's synaptic weights are randomly initialized and kept fixed, implying that the neuronal patterns within the reservoir remain unchanged after learning. The only trained part is the readout layer that connect the reservoir to the output. This fixed-weight approach do not model plasticity in the reservoir, but we argue that it is sufficient and consistent with György Buzsáki [2019]'s perspective. This is supported by experiments showing that learning only modestly and transiently impacts network dynamics [Golub et al., 2018]. Furthermore, this view of the navi-

gation system allows for a simple model that is easy to understand and manipulate, which is our main novelty and contribution compared to the other models listed: we don't require complex architectures or concepts to study splitter cells, just a pool of randomly interconnected neurons. Despite its simplicity, this model has already demonstrated great success and robustness in several works by Eric A. Antonelo [Antonelo et al., 2007, 2008, Antonelo and Schrauwen, 2009, 2012], and has even been successfully implemented on real robots navigating spatial environments [Aislan Antonelo and Schrauwen, 2015]. We demonstrate that this model can robustly solve a T-maze alternation task, specifically within a continuous state-space. Additionally, like the model in Raju et al. [2024], it handles ambiguous sensory inputs reflecting a more realistic scenario. Rather than relying on allocentric information such as spatial coordinates, the model has access only to partially observable inputs, i.e. egocentric information, like distance sensors. These sensors can read identical values in different locations within the maze, making the sensory information ambiguous. This feature increases the task's realism and difficulty and distinguishes our approach from Stachenfeld et al. [2017], Whittington et al. [2020], George et al. [2021], where such ambiguity was not explicitly addressed.

We show that the internal state of some units within the model exhibits similar characteristics to biological splitter cells. By conducting a theoretical analysis similar to that performed by Duvelle et al. [2023], we examined various hypotheses and their implications within the reservoir computing model. The results of our study demonstrate that the majority of these hypotheses are indeed validated by the reservoir computing model. While the model fails to solve the alternation task when splitter cells are removed, we show that re-optimizing the model without the splitter cells allows it to do the task again, with new splitter cells emerging. This strongly suggests that splitter cells arise from the behavior (the sequence of actions within the maze) rather than the other way around: it is not that the bot navigates because splitter cells exist. Additionally, we show that our model can decode place cells, head direction cells, and corner cells. Overall, this work aligns with previous studies but employs a simpler model consistent with György Buzsáki [2019]'s perspective.

1.2 Methods

1.2.1 Environment

Task presentation. The class of tasks called spatial alternation has been widely used to study hippocampal and working memory functions [Frank et al., 2000]. For the purpose of our investigation, we simulated a continuous versions of the task, wherein an agent must navigate through an 8-shaped track (see Figure 1.1) and alternate between right and left choices at a decision point, then returns to the central corridor, essentially following an 8-shape trace while moving. This 8-maze environment offers an ideal setup for observing splitter cells when the agent enters

the central corridor, because the agent can have various past trajectories leading to the central corridor and different future trajectories upon exiting it.

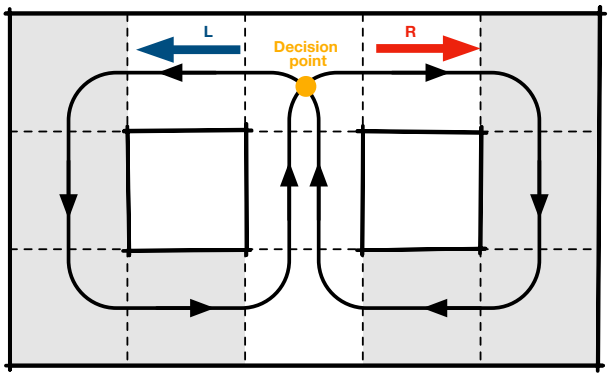

Figure 1.1: **An expanded view of a T-Maze**. At the decision point in yellow, the agent has to decide whether to go right (red) or left (blue).

Task variants. We examined two variants of the task. In the first variant, the agent is solely driven by the sensory input. In the second variant, the agent is helped with two contextual cues (labeled "L" for the left cue and "R" for the right cue), which are activated only within the central corridor and simulate visual signals that guide the agent's direction at the next intersection after traversing this corridor.

1.2.2 Modeling framework

Reservoir computing. Solving an alternating task necessitates the presence of a functional working memory, as the agent must retain information about its previous direction, such as turning right in the prior loop, to inform its decision to turn left at the subsequent decision point. Reservoir computing emerges as a promising solution for it, thanks to its fading memory [Pascanu and Jaeger, 2011]. This attribute enables the system to retain past experiences, facilitating their recall for subsequent decisions. Consequently, our model consists of a reservoir computing network of type Echo State Network (ESN) [Jaeger, 2007] that controls the movement of the agent solving a continuous navigation task in the 8-maze of Figure 1.1. An ESN is a recurrent neural network (called reservoir) composed of randomly connected units, associated with an input and an output layer. Inputs are projected into the reservoir, a non-linear, high-dimensional space, allowing the integration of information over time and space. Only the output neurons, referred as the readout neurons, are trained (in red in Figure 1.2). The neurons have the following dynamics:

$$\frac{1}{\alpha}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \tanh(W.\mathbf{x} + W_{in}.\mathbf{u}) \quad \text{and} \quad \mathbf{y} = W_{out}.\mathbf{x} \tag{1.1}$$

where $\mathbf{x}$, $\mathbf{u}$ and $\mathbf{y}$ represent the reservoir states, input, and output. W, W_{in}, and W_{out} are weight matrices, while *tanh* refers to the hyperbolic tangent function. α refers to the leak rate, a crucial parameter of the ESN that plays a central role in controlling the memory and timescale of the network's dynamics: a low leak rate indicates a longer memory and a slower dynamics, whereas a high leak rate leads to a shorter memory but a higher speed of update dynamics [Lukoševičius, 2012].

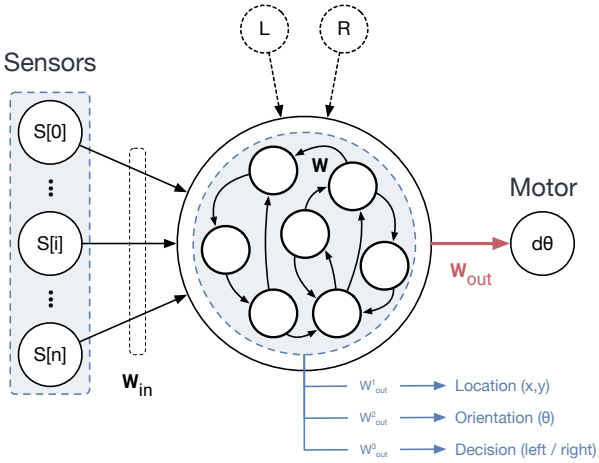

Figure 1.2: **Model Architecture**. Composed of 8 sensor inputs, a reservoir and a motor output (relative orientation change). The black arrows are fixed connections while the red arrows are plastic connections and are trained. A left (L) and right (R) cue can be fed to the model depending on the experiment. The reservoir states can be used to decode supplementary information such as location (x,y coordinates), orientation (absolute θ) and decision to go left or right.

TRAINING. Only the output weights W_{out} are trained, using a supervised learning. We tested two categories of training: offline learning and online learning rules. The offline learning rule uses linear ridge regression method (the Tikhonov regularization) on pre-generated data:

$$W_{out} = Y^{target} X^T (XX^T + \beta I)^{-1} \quad \text{with} \quad X = [1, \mathbf{u}, \mathbf{x}] \tag{1.2}$$

where Y^{target}, β and I are respectively the target signal to approximate, the regularization coefficient and the identity matrix. The online learning approach employs methods that enable task learning using only local temporal information, which is more biologically plausible than the global optimization of linear regression. Specifically, we utilized online Recursive Least Squares (RLS) for rapid weight updates. The readout weights are adjusted according to the following equation: It iteratively applies the following equations to train the readout of the networks:

$$\mathbf{W_{out}}(t) = \mathbf{W_{out}}(t - \Delta t) - e(t)\mathbf{P}(t)X(t) \tag{1.3}$$

With:

- $e(t)$ is the error at time step t: $e = Y^{target}(t) - Y(t)$
- $P(t)$ is the inverse correlation matrix at time step t: $\mathbf{P}(t) = \mathbf{P}(t - \Delta t) - \frac{\mathbf{P}(t-\Delta t)X(t)X^T(t)\mathbf{P}(t-\Delta t)}{1+X^T(t)\mathbf{P}(t-\Delta t)X(t)}$

This iterative process allows for the effective training of the network's readout by continuously updating the weights based on local error feedback.

The reservoir computing approach has been shown to exhibit some biologically plausible characteristics [Lukoševičius and Jaeger, 2009]. The model was built thanks to the python library ReservoirPy [Trouvain et al., 2020a].

In order to generate data for learning, we implemented a simple Braintenberg vehicle [Braitenberg, 1986] where the agent moves automatically with a constant speed and changes its orientation according to the values of its sensors. The model takes as input the sensor values $S_0, ..., S_n$ of the agent and outputs the next orientation θ (see Figure 1.2). At each time step the sensors measure the distance to the walls and the bot turns in order to avoid the walls. The sensors are limited in their ability to calculate distance, saturating for distances exceeding 100 distance units. Additionally, the sensor readings are affected by Gaussian noise, represented as $N(0, 0.05)$. A reading of 0 indicates proximity to an object, while a reading of 1 signifies that the object is far away or not detected. The sensors can record identical distance values even when the bot is not at the same spatial location in the maze, as illustrated in Figure 1.3. This phenomenon highlights the ambiguity of the sensory input. The challenge for the model lies in disambiguating this sensory information to accurately determine the bot's location. The decision to use distance sensors as input provides a simplified egocentric representation of the environment, abstracting the complexities of the multiple sensory modalities involved in visual and auditory processing that animals like rats typically utilize. However, we believe this simplification is justified, as the distance sensors still capture ambiguous sensory information. Furthermore, some animals, such as bats [Wenstrup and Portfors, 2011] and other species [Brinkløv et al., 2013, Evans, 1973], utilize echolocation, which offers distance information analogous to that provided by distance sensors.

At each timestep, the position of the bot is updated as follows:

$$\theta(t) = \theta(t-1) + 0.01 \sum_i \alpha_i s_i \tag{1.4}$$

$$p(t) = p(t-1) + 2 * (cos(\theta(t)) + sin(\theta(t))) \tag{1.5}$$

where $p(t)$ and $p(t+1)$ are the positions of the agent at time step t and $t+1$, $\theta(t)$ is the orientation of the agent, calculated as the weighted (α_i) sum of the values of the sensors s_i. The norm of the movement is kept constant and fixed at 2.

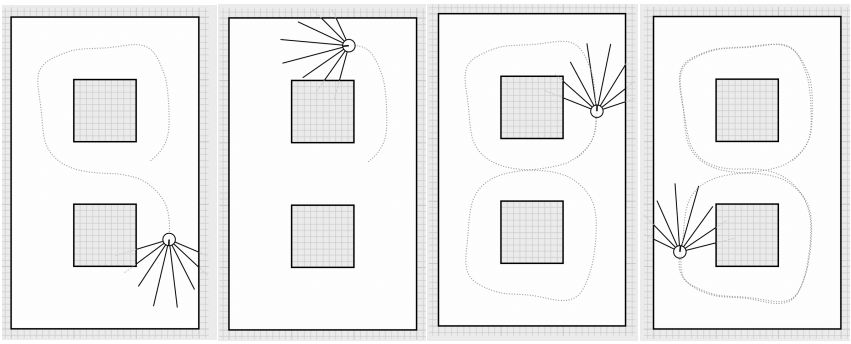

Figure 1.3: **Ambiguous inputs during the agent simulation**. Snapshots from a simulation showcasing the agent's navigation within an 8-maze. These snapshots illustrate situations where the sensors record the same distance values, even though the bot is not positioned identically in the environment. This highlights the reservoir model's challenge of handling ambiguous sensory information.

The ESN is trained using supervised learning, containing samples from the desired 8-shaped trajectory. Since the Braitenberg algorithm only aims at avoiding obstacles, the agent is forced to follow the desired trajectory thanks to the timely addition of walls at the intersection points as shown on Figure 1.4. After generating the right pathway, the added walls are removed and the true sensor values are gathered as input. Gaussian noise is introduced to the robot's sensor readings at each time step to enhance the robustness of the training process. Approximately 10,000 time steps were generated, equivalent to 14 complete alternating 8-loops, and divided into 75% for training and 25% for testing. The training and testing processes completed in just a couple of seconds, making the simulation process fast and easy to use, modify, implement, and optimize.

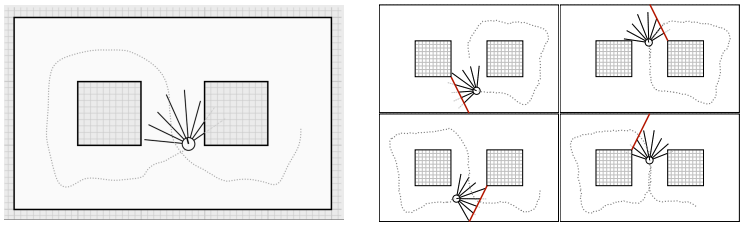

Figure 1.4: **Generation of the 8-shape trajectory** with the addition of walls at the intersection points

Modeling contextual cues. In the first task variant without the contextual cues, the reservoir model is solely driven by the sensory input from the bot. For the second task variant with contextual cues, the model incorporates two additional inputs, referring to the "contextual inputs" (labeled "L" for the left-turn cue and "R"

for the right-turn cue in Figure 1.2). They are binary values, set to 1 only when the agent is expected to turn in the corresponding direction.

1.2.3 Analysis

Task solving. After the training phase, we look at the model's ability to solve the task by directly looking at the online simulation. If the bot can perform the task and execute at least two loops (one loop is composed of 700 time steps with one alternation, going one time to the right and one time to the left), the model is considered to have successfully solved the task. When the agent navigates the maze, the reservoir's internal states are recorded and analyzed using two methods: single-cell level analysis and population-level analysis.

Single-cell level analysis. We analyze the mean firing activity of individual neurons within the reservoir to observe whether we can observe various types of hippocampal cells including head-direction cells, place cells, and splitter cells. The differentiation in mean firing activity on intersecting trajectories serves as a qualitative indicator of splitter cell behavior. To quantitatively assess individual splitter cells, we used one of the detection method cited in Duvelle et al. [2023]-Box 3. We use ANOVA (Analysis of Variance) method used to detect splitter cells in hippocampal studies [Wood et al., 2000]. It is a statistical technique to assess whether there are significant differences in neuronal firing rates across different behavioral conditions. We record the firing activity of individual neurons in the reservoir as the agent traverses the central stem of the T-maze, which corresponds to a shared portion of different trajectories. The firing rates are grouped based on the two different conditions: right to left (RL) or left to right (LR). We recorded at least 3 activities per trajectory. The resulting F-statistic and p-value determine if there's a statistically significant difference between the trajectories. If a neuron shows a statistically significant difference in firing rates between conditions with $p < 0.01$, it is classified as a splitter cell. We visualize the detected splitter cells by generating their raster plot. Their activity are recorded in the central corridor during overlapping trajectories, subsequently turned into a dynamic firing rate according to the Poisson process [Heeger et al., 2000].

$$P\{1 \text{ spike in } \delta t\} \approx r\delta t \quad \text{and} \quad r(t) = \alpha[V(t) - V_{th}] \tag{1.6}$$

where α is the maximum firing rate 150 Hz, $V(t)$ is the normalized activity of the reservoir neuron and V_{th} is the mean normalized activity of the neuron.

Population-level analysis. We analyze the reservoir state of the ESN at the population level using Principal Component Analysis (PCA) as a dimensionality reduction technique to identify patterns and significant features within the processed

data. Notably, it is possible to observe the splitting activity inside the neural trajectories, as depicted in Figure 1.5.

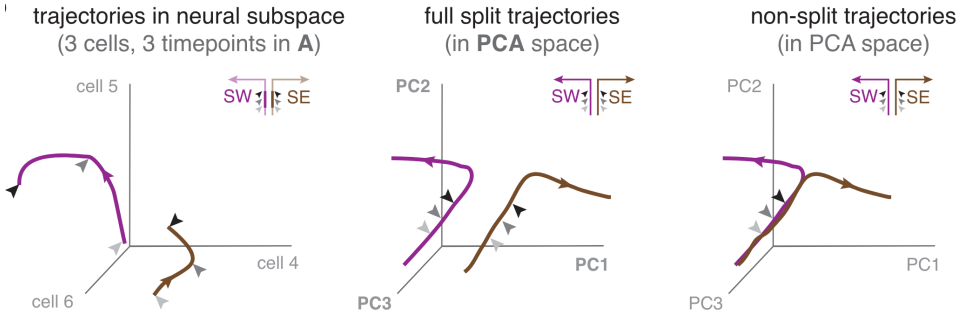

Figure 1.5: **Hypothesized neural activity trajectories of multiple splitter cells in the neural space**. In the middle graph, the SW and SE trajectories have distinct areas of neural activity space, even though the agent is going through the same locations, indicating a 'splitting' effect. In the case where there is no splitting effect, the PCA looks like the right graph, depicting overlapping neural trajectories. The Figure is taken from Duvelle et al. [2023].

Lesioning splitter cells within the model. We investigated the effects of lesioning the splitter cells within the model. To do this, we first identify the splitter cells and subsequently inactivate the connections associated with them by setting their weight values to 0 in W_{in}, W, and W_{bias}. This process effectively rendered the splitter cells silent. The resulting model, denoted as the *Lesioned model*, is not always able to successfully perform the alternating task and requires some adjustments:

- We bypass its disability of solving the task by feeding it with the exact sequence of sensory inputs the bot would receive if it were correctly performing the alternation task, without using the output model to compute the next position. This version of the model is referred as the *Forced Lesioned model*.
- We re-optimize the *Lesioned model* so as to make it performing the alternating task again (without modifying its connectivity structure). This version of the model referred as the *Optimized Lesioned model*.

The *Lesioned*, the *Forced Lesioned* and the *Optimized Lesioned* models possess the same connectivity structure. We subsequently analyse their internal activity when executing the task.

Decoding other hippocampal cells. In addition to splitter cells, we explore whether the reservoir's dynamical systems could encode other hippocampal cell-like activities, such as head-direction cells, decision cells, and place cells. Since these

cells are not explicitly represented in the reservoir, we employ an additional readout, referred as the "decoder readout", trained using offline supervised learning to decode their activity. The decoder was trained on 7500 time steps and tested on 2500 time steps, with a regularization factor of 1e-3 for all readouts. For place cells, we generate training data by dividing the maze into 60 grid squares (10 by 6). Each square is labeled as 1 if the bot is located within it, otherwise it is set to 0. Head-direction cells, which fire in response to a specific orientation, are decoded by generating output vectors representing five distinct orientations: $[-\pi/2, -\pi/4, 0, \pi/4, \pi/2]$. For example, if the bot's orientation is $-\pi/2$, the output vector will be [1, 0, 0, 0, 0]. For decision cells, we generate training data by recording the bot's next decision (right or left) at the intersection following the central corridor, labeling these decisions as 0 (left) or 1 (right). These decoder readouts allow us to assess whether this spatial information is encoded within the reservoir dynamics. To complement this analysis, we also examine whether this spatial information could be directly decoded from the sensory inputs alone. To do so, we employed a Support Vector Machine (SVM) classifier [Cortes and Vapnik, 1995] to categorize the sensory inputs, evaluating whether it was possible to predict the robot's position, orientation, or decision based solely on the sensor data.

1.3 Results

1.3.1 Model performances

Robustness. Using an architecture composed of fixed recurrently connected neurons (the reservoir) and a trainable readout layer trained with linear regression, both models, with and without contextual cues as input, were able to successfully perform the alternating task for at least two loops. Both models were successfully trained using two different training rules, offline regression and online RLS learning, demonstrating the flexibility of the learning methods. The optimized hyperparameters of the model are presented in Table of the Supplementary Data (Section 1.5). However, the model is quite robust to variations in these hyperparameters. This robustness was also demonstrated in the work of Aislan Antonelo and Schrauwen [2015]. While we fixed the number of units to 1000, a model with 500 neurons also performed well, provided the other hyperparameters were optimized accordingly. The values for input and output connectivity could also vary across a wide range. The leak rate and spectral radius were among the most critical parameters, requiring joint optimization to manage the level of chaoticity, as well as the timescale and memory of the model. The regularization parameter, which helps reduce overfitting, was another important aspect to optimize. All parameters were manually optimized, and the optimal settings were quite straightforward to identify. Overall, we present a set of hyperparameters that yielded good performance, but many other combinations also worked, underscoring the robustness of our models. Due to the slower pace of online learning, we opted to conduct the rest of the analysis using the offline learning rule.

Influence of inner structure. We subsequently evaluated the extent to which the architecture of the reservoir model influenced its ability to solve the alternating task. To achieve this, we generated 15 random connectivity matrices W and tested them using the same set of hyperparameters, including the leak rate, regularization parameters, and input and reservoir connectivity. The matrices W were generated with a fixed connectivity, where the non-zero elements were drawn from a normal distribution $\mathcal{N}(0,1)$. Each matrix was then scaled by a factor of $\frac{\text{sr}}{\rho(\mathbf{M})}$ to ensure the matrix has the desired spectral radius as shown in equation 1.7.

$$W = \frac{\text{sr}}{\rho(\mathbf{M})}\mathbf{M}, \quad \text{where} \quad \mathbf{M}_{ij} = \mathcal{N}(0,1) \quad if \quad non-zero \quad element \tag{1.7}$$

All connectivity matrices W were effective in solving the alternating task, with no changes to any hyperparameters other than the spectral radius. This result highlights the model's robustness to different architectures, indicating that its success did not rely heavily on the specific architecture, apart from the requirement for some level of recurrence.

1.3.2 Splitter cells

1.3.3 Splitting effect at the single cell level

In their recent review, Duvelle et al. [2023] provided a comprehensive analysis of splitter cells, examining both experimental and theoretical aspects. Our goal is to replicate this analysis and verify whether their hypotheses hold true within the reservoir model. Using the ANOVA method, we identified 281 splitter cells that exhibited significantly different firing rates in the central stem ($p < 0.01$), even when the bot received identical inputs in both cases. This difference in firing activity is illustrated in Figure 1.6, which shows neuronal activity recorded as the bot approaches the central corridor (indicated as the grey zone in the figure). Each trace corresponds to the firing activity of a splitter cell in the central corridor over 50 time steps. We recorded two sequences of traces as the bot completed two consecutive loops in the central stem (each loop consisting of alternating movements to the right and left). The differences in firing activity are evident, with red traces representing activity when the bot is moving from left to right and blue traces when moving from right to left. Raster plots of the same splitter cells are generated from the Poisson process and depicted in Figure 1.7, illustrating the average firing activity of a subset of splitter neurons when the bot is traversing the central corridor. Various patterns are evident: certain neurons, such as those with indices 257, exhibit differences in firing rates, others neurons fire exclusively for specific trajectories, resembling the behavior of the neuron indexed as 920. Additionally, some others such as those indexed as 275 and 75, fire at different timings. Each individual neuron displays a

distinct average activity pattern based on the trajectory, even when recorded at the same spatial location.

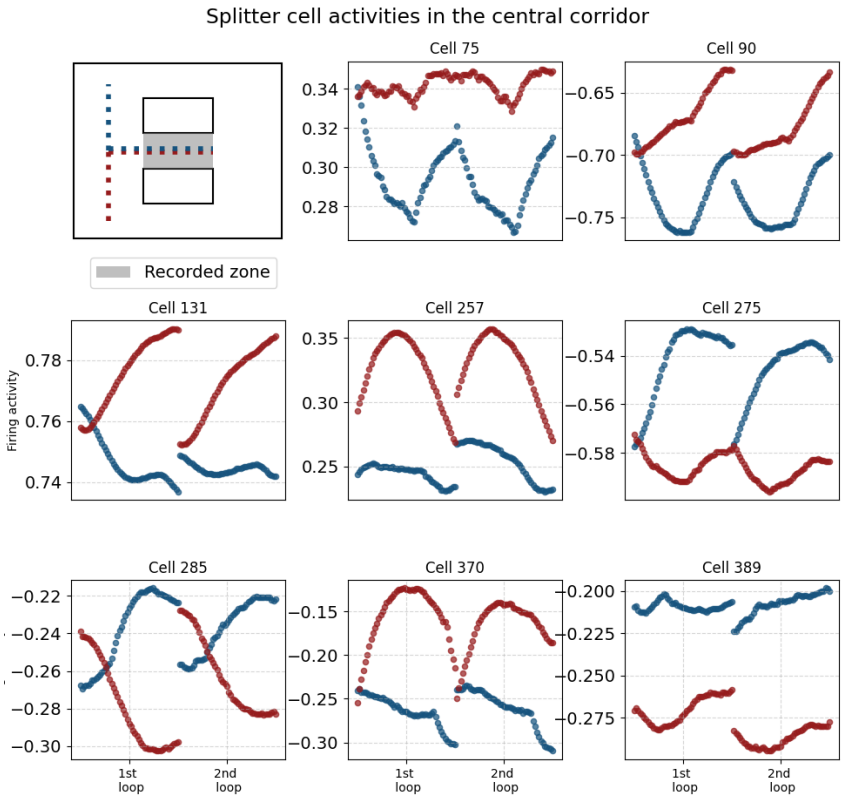

Figure 1.6: **Comparing firing rates of 8 splitter cells in the central stem**. The identification of these cells was achieved using the ANOVA method, which revealed significantly different firing activities ($p < 0.01$) in overlapping trajectories 'RL' and 'LR'. The activity of splitter cells is recorded when the bot enters the central corridor (indicated as the grey zone in the bottom right figure). Each trace represents the firing activity of a splitter cell in the central corridor over 50 time steps. Two sequences of traces correspond to the bot completing two consecutive loops, thus going the central stem 4 times (one loop consists of going 2 times in the central stem, alternating right and left movement). The differences in firing activity are evident: traces are colored red for activity when the bot is moving from left to right and blue when moving from right to left.

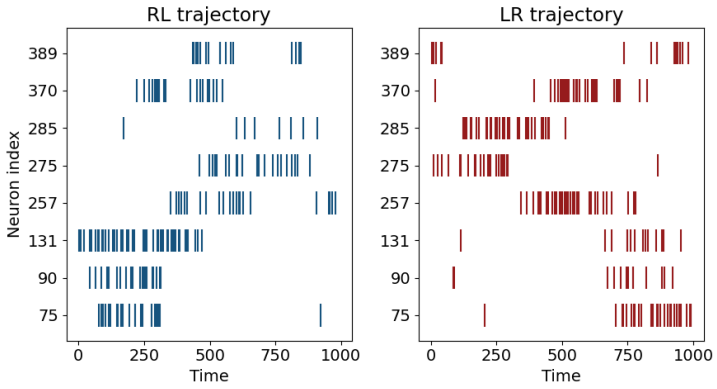

Figure 1.7: **Raster plots**. Different types of pattern are depicted: the same neurons can have a difference in firing rate, or fire for only one of the trajectories. The different colors correspond to the trajectories: R-L (blue) and L-R (red).

1.3.4 Splitting effect at the population level

Temporal Context Model versus Latent State Inference Model The splitting effect is observed at the population level by analyzing trajectories in the neural space. Splitter activity is characterized by distinct neural trajectories that emerge when the animal is at the same spatial location. Figure 1.8 from Duvelle et al. [2023], illustrates these neural trajectories. In Figures 1.8-C and 1.8-D, the authors present hypothesized neural activity trajectories during the execution of an alternation task. Two trajectories are shown: a blue one representing mouvement from the left loop to the central corridor, and a red one showing mouvement from the right loop to the same central corridor. Despite sharing the same physical location when the animal traverses the central corridor (where the markers C1, C2, and C3 are placed), these trajectories occupy distinct regions in PCA space. Without the splitting effect, the neural representations of these three markers would overlap, as they correspond to the same spatial location. However, they are distinctly separated with a measurable PCA distance, indicating the presence of differentiated neural activity. Duvelle et al. [2023] argue that this observation provides evidence for the presence of splitter cells in the neural population. The distinct Figures (1.8-C and 1.8-D) correspond to two different models of splitter cells. Figure 1.8-C depicts the "Temporal Context Model" (TCM), which is sensitive to prior experiences. This means that the current neural state of the animal varies depending on its past. When the recent past is similar, population activity at a given location in the maze exhibits closer proximity in PCA space. Conversely, when the recent past differs, the distances increase. This phenomenon is demonstrated by the varying distances between the markers C1, C2, and C3, each representing specific positions within the central corridor. When the animal reaches C3 in the maze, it has traversed the entire central corridor, resulting in a shared trajectory that accounts for the smaller distance between the C3 markers. In contrast, the animal's past experiences at C1

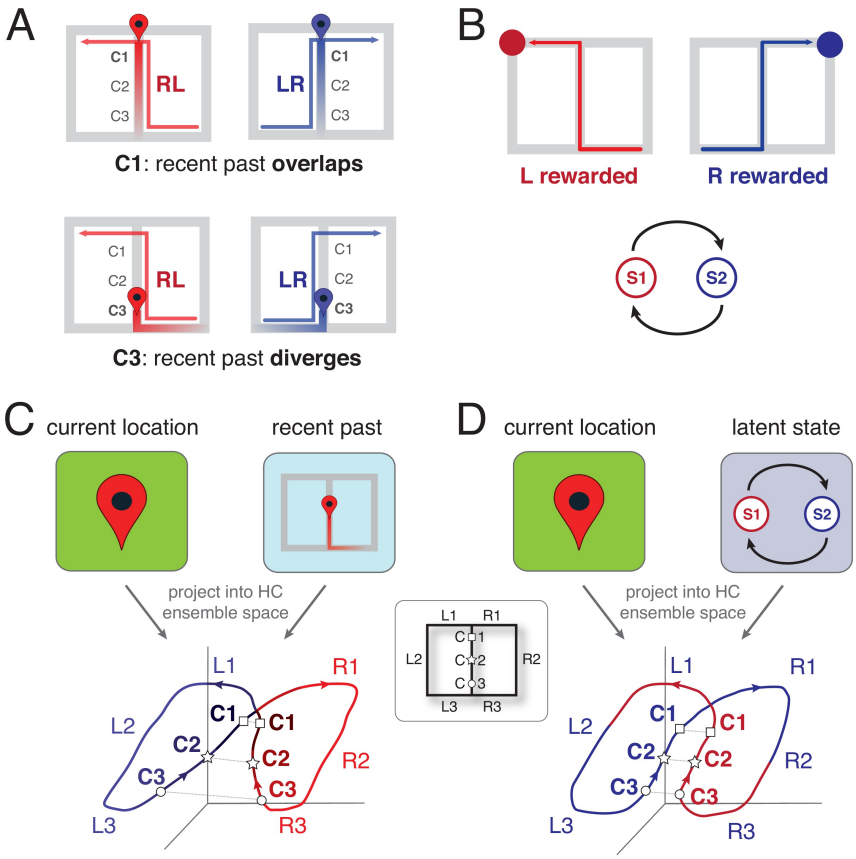

Figure 1.8: **Temporal context model (left) vs. Latent State Inference (right) Left**: hypothesized neural Activity Trajectories in PCA Space illustrating the temporal context model (TCM). The population activity at a specific location exhibits closer proximity in PCA space when the recent past is similar and greater distance when it is less similar. This is demonstrated in the distances between markers C1, C2, and C3. **Right**: hypothesized neural activity trajectories for the latent-state inference (LSI) hypothesis. The distances between the markers C1, C2, and C3 don't vary. The Figure is taken from Duvelle et al. [2023].

differ significantly, which explains the greater distance between the C1 markers. Figure 1.8-D illustrates the "Latent State Inference" (LSI) model. Unlike the TCM, the LSI model organizes experiences into discrete states, resulting in a neural state less sensitive to historical context. As a result, the distances between the two neural trajectories in the central corridor remain constant: the distance between C1 markers is equal to that between C2 markers, and the distance between C3 markers is also consistent. The Latent State Inference (LSI) model is suitable in situations where experiences can be categorized into distinct states without strong dependencies on past experiences. For example, in scenarios where each stimulus is associated with a specific state, the categorization remains valid regardless of how previous stimuli were encountered. Duvelle et al. [2023] illustrates this with the example of one arm being rewarded while the other is not.

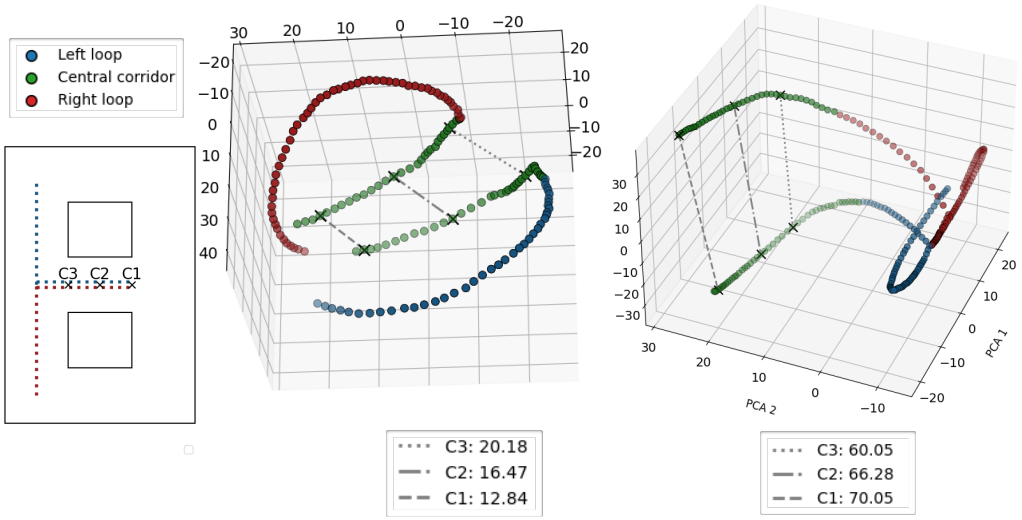

Figure 1.9: **3D PCA of the reservoir states**. **Left**: PCA applied to the model without contextual cues, resembling the TCM model introduced in Figure 1.8. The increasing distances between C1, C2, and C3 markers suggest an amplified influence of past trajectories on neural activity. **Right**: PCA applied to the model with contextual cues (leak rate$\alpha = 0.1$), similar to the LSI model introduced in Figure 1.8. The decreasing distances between C1, C2, and C3 markers indicate more distinct neural activity between the two trajectories, reflecting a stronger association with one of the two discrete states.

Implicit approximation of the state-space environment and splitting effect at the population level We proceed to a similar population analysis by applying 3D PCA to the reservoir states. This analysis was conducted for both the model with contextual cues (Figure 3.8-right) and without contextual cues

(Figure 3.8-left). In Figure 3.8, the blue trace followed by the green trace indicates the agent traversing the left loop and heading to the the central corridor, while the red trace followed by the green trace indicates the agent traversing the right loop and heading to the central corridor. These two trajectories occupy distinct regions in the PCA space, despite they share the same location when the bot is traversing the main corridor (green trace). The three markers (C1, C2, C3) indicate specific positions within the main corridor. These three markers are distinctly separated with a measurable PCA distance. This observation demonstrate the presence of splitter cells. In both case PCAs show two linearly separable sub-attractors, corresponding to the two loops of the 8-shape. This result demonstrates the ESN's ability to implicitly approximate the state-space of the 8-maze environment through continuous temporal dynamics. The model achieves this by mapping the input data into a higher-dimensional space[Jaeger, 2001]. The recurrent connections within the reservoir enable the network to incorporate and reverberate past information, allowing for the encoding of temporal context and the integration of the past trajectory.

The model without contextual cues corresponds the TCM. In the model without contextual cues, we observed decreasing distances between C1,C2 and C3 markers in the central corridor (Figure 3.8-left), which aligns with the TCM (Figure 1.8-C). This makes sense, as the model is influenced not only by current sensory inputs but also by recently experienced events, as the past trajectory of the bot is stored through the reverberation of recurrently connected neurons, allowing for a context-dependent response to present events.

The model with contextual cues corresponds the LSI. In the model with contextual cues, we observed the opposite effect: as the bot enters the central corridor, the distances between neural trajectories increase, indicating more distinct neural activity and a stronger association with one of the two discrete states (Right or Left). We conducted this analysis with varying values of the leak rate, which corresponds to the hyperparameter that controls sensitivity to past activities. A low leak rate results in greater memory capacity, making the model more sensitive to previous information, while a high leak rate indicates smaller memory capacity and reduced sensitivity to past experiences. Figure 3.8-left displays activity for a leak rate of 0.1, which reflects high past dependencies; nonetheless, we still observed a more discrete configuration consistent with the LSI model (Figure 1.8-D). Furthermore, we applied the same PCA analysis to a model with cues and a leak rate of 1 (indicating no memory). The resulting PCA, shown in Figure 1.10, reveals consistent distances between neural trajectories in the central corridor (Figure 3.8-right), also aligning with the LSI model. The presence of contextual cues provides direction information at each time step within the central corridor, positioning the model in

a more discrete configuration. Consequently, the model does not require extensive memory of the past trajectory to successfully complete the navigation task.

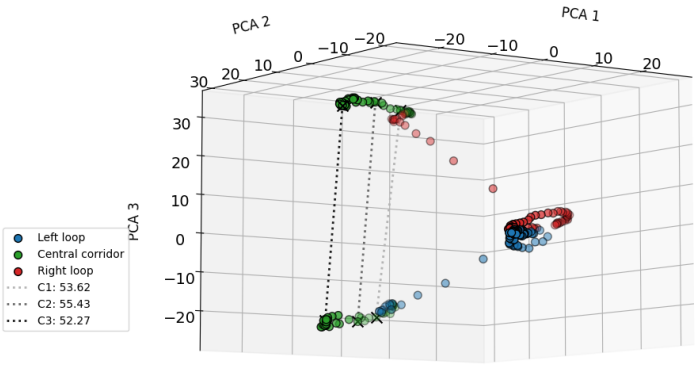

Figure 1.10: **3D PCA of Reservoir States applied to the model with contextual cues (leak rate = 1)**. The dotted lines represent the distances between markers C1, C2, and C3, which remain constant as the agent enters the central corridor. This behavior parallels the dynamics observed in the LSI model shown in Figure 1.8-D.

These results confirm Duvelle et al. [2023]'s theoretical framework introducing TCM and LSI. These two theories aren't mutually exclusive, leaving possibility for hybrid models that contain elements of both. While reservoir computing primarily relies on maintaining working memory within the reverberation of reservoir neurons, it's natural to view this model as having a temporal gradedness comparable to TCM. This forms the basis of reservoir computing. By incorporating contextual cues as input, it is possible to transition from a TCM to LSI model. Adjusting the leak rate of the models allows for a flexible hybrid model, combining temporally decaying representations of the past with discrete state representations. For instance, in tasks like the continuous alternation task, where task-state switches occur with changes in cue values, the internal states shift between distinct regions of activity space while still permitting continuous neural activity.

1.3.5 Silencing splitter cells inside the models

Model without cues. We identified 49 splitter cells (SC1) in the initial model without contextual cues. These SC1 cells were subsequently silenced, and the model re-optimized by adjusting only the leak rate and spectral radius, without altering the structural connectivity. Adjusting the leak rate and spectral radius compensated for the absence of splitter cells by introducing more chaos and inertia into the reservoir dynamics, allowing the model to maintain task performance. Despite the inactivation of SC1, the *Optimized Lesioned model* successfully performed the alternation task. Notably, 7 new splitter cells (SC2), distinct from SC1, emerged. Five

of these SC2 cells are shown in Figure 1.11, column 'Optimized Lesioned model', where their splitting activity is visible as the bot traverses the central corridor from different trajectories (red for right, blue for left). In contrast, these same neurons do not exhibit splitting activity in the Initial model (see Figure 1.11, column 'Initial model'), confirming their absence from the SC1 population. This transformation demonstrates that silencing SC1 led to a reorganization of the model, giving rise to new splitter cells (SC2) in the *Optimized Lesioned model*, which were not initially splitter cells. Without optimization, the lesioned model cannot perform the task independently. To address this, we forced the non-optimized lesioned model to execute the task by providing it with the exact sensory inputs the bot would receive if it were correctly performing the alternation task. Interestingly, forcing the model to alternate still resulted in the emergence of 638 splitter cells, as shown in Figure 1.11, column 'Forced Lesioned model'. Although some neurons from SC2 in the Optimized Lesioned model did not show splitting activity in the Forced Lesioned model, others did, highlighting that the identity of splitter cells does not always overlap across models. The splitting activity in both the Optimized and Forced Lesioned models is also evident at the population level when applying 3D PCA to the internal states of the reservoir, as shown in Figure 1.12. The neural trajectories in the central corridor (in green) do not overlap, illustrating this effect. These results support that as long as the model does the alternation task, splitter cells will emerge, potentially from different neuronal populations. Splitter cells represent a dynamic neural mechanism that is observable during task execution. Performing a specific task activates distinct neural patterns, regardless of the model's optimization state, suggesting that splitter cells arise as a direct consequence of executing the correct sequence of actions.

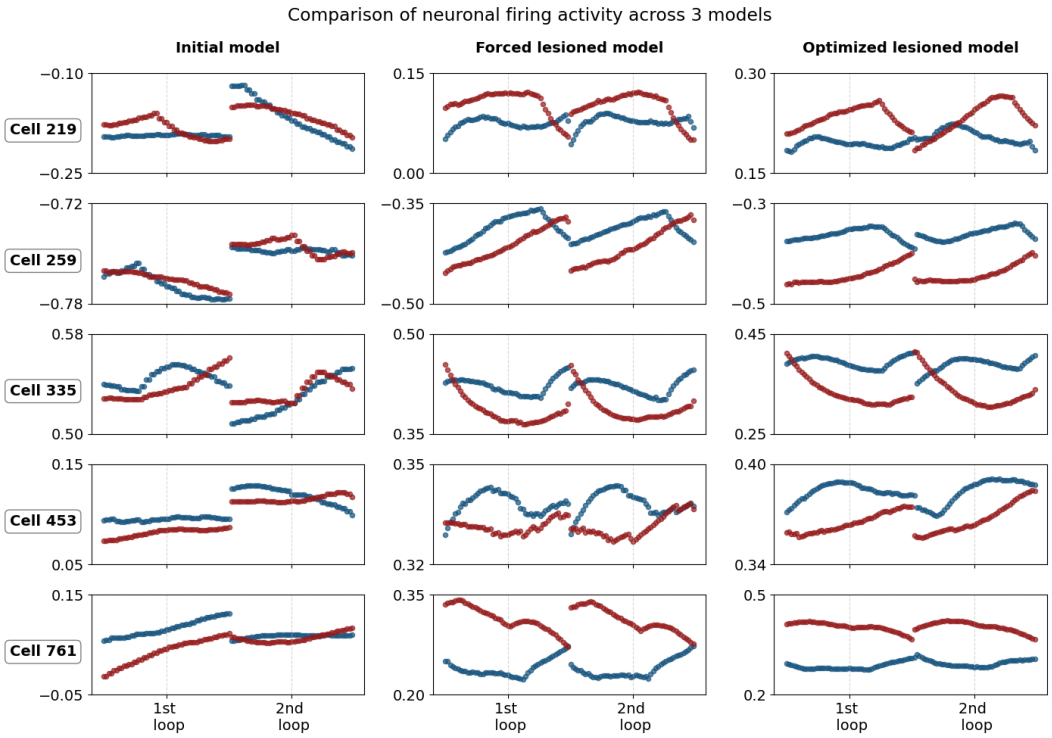

Figure 1.11: **Firing activity of SC2 in the initial model, the lesioned model force to do the task and the optimized lesionned model.**

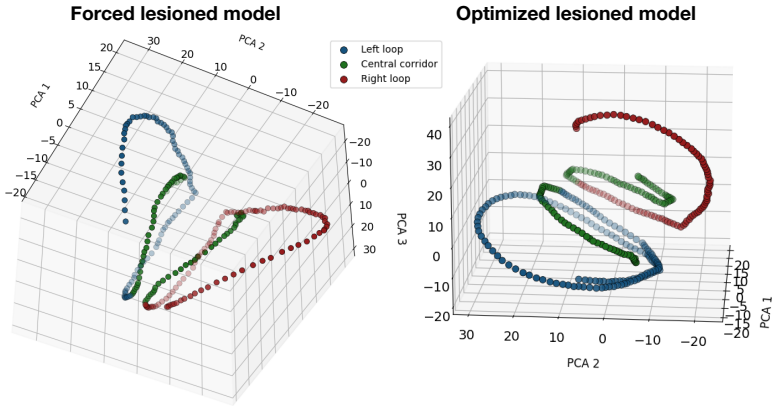

Figure 1.12: **3D PCA of Reservoir States applied to the forced lesionned and optomized lesion models.**

Model with cues. We identified 59 splitter cells (SC1) in the initial model with cues, which were subsequently silenced. Despite this inactivation, the Lesioned model (without re-optimization) still successfully performed the alternation task. Remarkably, 15 new splitter cells (SC2), distinct from SC1, emerged. These SC2 cells were also silenced, resulting in a model lesioned twice, yet it still performed the task successfully. We repeated this process three times, each time observing the emergence of new splitter cells that had not exhibited splitter activity in the previous lesioned version. Specifically, the initial model contained 1,000 active neurons with 59 splitter cells; the second model, after SC1 inactivation, had 941 active neurons (1,000 - 59) and 15 new splitter cells; the third model had 926 active neurons (941 - 15) with 68 new splitter cells; and the fourth model had 858 active neurons (926 - 68) with 34 new splitter cells. All models were identical in structure, and no re-optimization was required, unlike the model without cues. This is likely due to the contextual cues informing the model of the correct decision at each time step in the central corridor, making the task less challenging. Even with the removal of splitter cells, the model with cues had sufficient information within its dynamics to continue making the correct decisions. The repeated apparition of new splitter cells after silencing the previous ones strongly reinforces the idea that splitter cells are a direct consequence of task execution. The splitter cells are induced by the behavior itself.

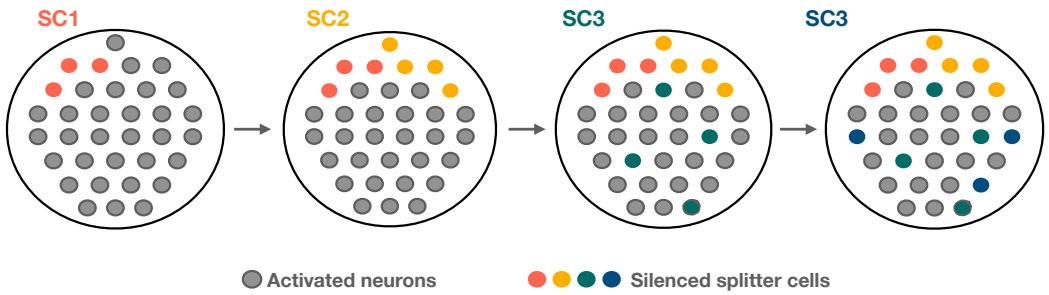

Figure 1.13: **Effect of silencing splitter cells.** Colored cells represent the lesioned splitter cells, while grey cells indicate active neurons. The network connectivity remains unchanged

1.3.6 Decoding other hippocampal cells

We added a decoder readout to the reservoir to extract additional spatial information as the bot navigated, referred as the readout decoder.

Place cell-like activities. The decoder readout achieved a success rate of 96%, correctly identifying the bot's location. While navigating, the decoder consistently activated the correct square corresponding to the bot's position, mimicking the behavior of place cells. This result indicates that place cells can be decoded from the reservoir states. Training the decoder readout to directly decode the position of the bot was also successful as depicted in Figure 1.14-right. These results demonstrate that place-cells-like activity and spatial coordinates of the bot can be decoded from the reservoir thanks to its capacity of implicitly approximate the state-space of the environment.

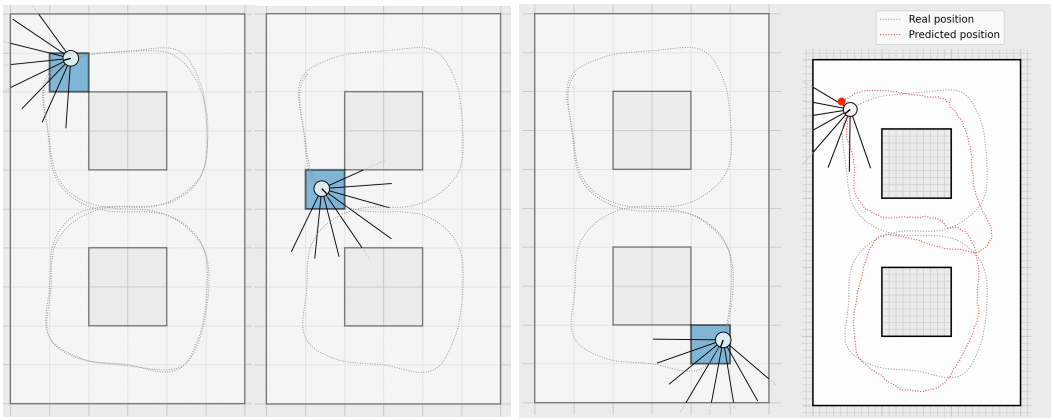

Figure 1.14: **Snapshot of the simulation during bot navigation and place cell decoder predictions. Left:** Blue squares indicate place-cell-like hippocampal activity, successfully decoded by the readout decoder. **Right:** The grey dotted line represents the actual trajectory of the bot, while the red dotted line shows the predicted bot position at each time step. The large red dot marks the current prediction. The decoder readout accurately tracks the bot's position within the maze.

Decision cell-like activities. The decoder readout achieved a 100% success rate in predicting the bot's next decision. As shown in Figure 1.15, when the bot recently made a right turn, the reservoir adjusted its internal state to favor a left turn within a few dozen time steps. This indicates that simply by observing the reservoir's internal dynamics, it is possible to anticipate the bot's next move. This finding suggests that the internal neurons of the model encode information about future decisions.

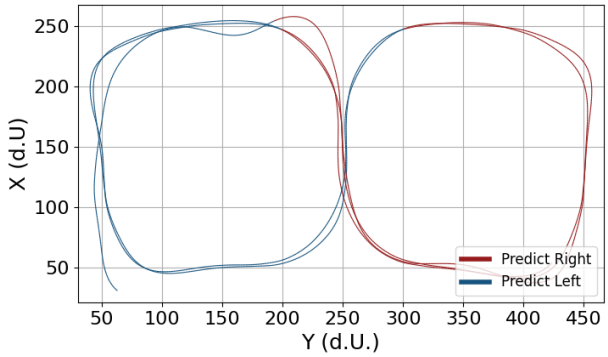

Figure 1.15: **Decoder predictions.** Agent's Trajectory and decision classification. Colors represent decoder predictions: red for right, blue for left directions. By simply observing the internal states of the reservoir, it is possible to predict the next prediction: the neurons of the agent encapsulate information about future predictions.

Head-Direction cell-like activities. The decoder readout achieved a 98% success rate in predicting the bot's orientation. This result demonstrates that the bot's head direction can be reliably decoded from the reservoir's internal states.

Sensor-Based Decoding. In contrast, applying the same decoding techniques directly to raw sensor inputs failed to predict place cells, head-direction cells, or decision cells, as expected. This confirms that the reservoir's internal states encode critical information for these cell-like activities, while the ambiguous sensory data alone does not.

1.4 Discussion

The model we proposed to study is as simple as it can be: a random recurrent network is fed by sensors that allows to compute a direction change. Despite this apparent simplicity, the model is able to solve a continuous and non-trivial alternating navigation task, using ambiguous, limited and egocentric sensory information. Inside that model, we've been able to demonstrate for the existence of splitter cells as described by Duvelle et al. [2023] in their review. By redoing their analyses on the reservoir, we've found activities that characterize splitter cells in rodents and show how they re-emerge following a targeted lesion and retraining. Furthermore, we've also validated the proposed theoretical framework by developing a model that incorporates concepts from both TCM (Temporal Context Model, Howard and Kahana [2002]) and LSI (Latent State Inference, Gershman and Niv [2010]). These concepts, that appear to be two distinct notions, are actually both needed for the interpretation of experimental data on splitter cells. The model we proposed suggests a simple mechanism where the transition from TCM to LSI is driven by contextual cues.

Even though we did not claim any architectural plausibility with the hippocampal formation, the structure of the model is reminiscent of the CA3 structure with its highly recurrent nature, with the notable absence of learning inside the reservoir. In our case, the learning process is a simple linear identification of a random pattern of activity inside the reservoir in order to issue the proper motor command. This process is indeed well aligned with György Buzsáki [2019]'s inside-out view of the brain, where he suggests that brain activity is not generated from scratch during exploration and learning. According to this view, the brain can be considered as a vast reservoir capable of generating an extensive repertoire of neuronal patterns, initially independent of experience. In other words, these patterns have initially no meaning and acquire it through exploration and learning. In this sense, experience is primarily a process of matching preexisting neuronal dynamics to outside events, and learning does not create new brain activity from scratch, but rather selects preexisting pattern of activity responding to the external stimuli. This "inside-out" view of brain function is further reinforced by findings showing that spatial memory cells and their properties are already present when young pups first begin to move [Langston et al., 2010], or even emerge early in postnatal development, before significant spatial exploration takes place [Muessig et al., 2016, Wills et al., 2014]. These studies indicate that the hippocampus is capable of generating a wide range of possible neuronal trajectories even before the organism begins exploring its environment. On this regards, it is interesting to note that the randomness of the network in our model is not strongly constrained such that any matrix with the proper spectral radius will provide a sufficiently rich reservoir of dynamics from which the behavior can be acquired. Furthermore, in case of targeted lesion, it is relatively easy for the model to re-acquire a behavior based on the new dynamic inside the reservoir, leading to the emergence of new splitter cells.

Beyond splitter cells, we've also decoded the presence of place cell-like, head direction cell-like and decision cell-like activity units. Even though we did not look for each and every other type of cells, we're confident that we would probably be able to decode them. This strongly suggests that their activities do not derive from structure nor learning (we have no learning, and no feedback inside the reservoir) but are rather indicative of a correlation with some pre-existing dynamics that are linked to the actual behavior. Considering for example decision cells (indicating decision to go left or right), it comes as no surprise that we can identify them inside the reservoir since the agent is really going left or right at the end of the central corridor, at least for an external observer. For the agent however, the sensory world is fully described by a set of eight sensors that do not convey the notions of *left* or *right*. Ultimately, this means that these decision cells may only exist in the eye of the observer. In other words, if a reservoir can solve a task with a minimal set of inputs, any variable that can be decoded from the activity of the reservoir derived from the resolution of the task itself as opposed to the specificity of a structures involved in the resolution.

During the past decade, the understanding of neural coding has gradually shifted from a population vector sum were a behavioral output is generated when the sum exceed a threshold, to a more complex theory that involve multidimentional neural activity space [Churchland and Shenoy, 2024]. This theoretical framework implies several subspaces: a null space, that refers to the subspace of neural activity that does not directly affect certain output variables, and a task-relevant subspace. In this context, splitter cells (and any other task related cells) reflect only a small partial dimension of the global encoding of the neuronal population. Disturbing the whole multidimentional space by lesioning/silenting specific neurons induces a general reorganization of the neuronal activity that affects the neural response of the remaining neurons to the task, hence "creating" new splitter cells, place cells, etc. This reorganization may, with some training, allow the network to adapt and to reproduce the same behavioral output as before. If we demonstrated how this can be achieved using supervised offline and online training inside the model, the question of continuous and gradual adaptation remains open.

Finally, in this chapter, we demonstrated that an unstructured neural networks can solve a complex decision-making tasks, even when faced with a continuous state-space environment and ambiguous sensory information presented to the agent. Additionally, our analysis revealed similar neural properties as those observed in the hippocampus, suggesting that the observed spatial properties in the network appear to rather be a byproduct of the sequence of actions taken by the agent, arising from the network's attempt to solve the decision-making task.

1.5 Supplementary data

1.5.1 Code availability

The code associated with this work is publicly available on GitHub at https://github.com/naomichx/splitter_cells_in_reservoir

1.5.2 Models parameters

Offline linear regression learning

These parameters represent a set of parameters that are making the models work, however these values can change and vary since the model is robust in changing the hyperparameter. This is particularly the case for the model with cues. In this case, the leak rate can vary between 0.08 to 1, the spectral radius can vary between 0.99 to 1.4.

Table 1.1: Initial model (offline learning)

Parameter	No cues	With cues
Input size	8	10
Number of units	1000	1000
Input connectivity	0.2	0.2
Reservoir connectivity	0.1	0.1
Input scaling	1	1
Spectral Radius	1.05	1.4
Leak Rate	0.1	0.2
Regularization	1e-5	1e-3

Table 1.2: Optimized Lesioned model

Parameter	No cues	With cues
Input size	8	10
Lesioned units	49	933
Number of units	951	67
Input connectivity	0.2	0.2
Reservoir connectivity	0.1	0.1
Input scaling	1	1
Spectral Radius	1.05	1.4
Leak Rate	0.07	0.09
Regularization	1e-5	1e-3

Online RLS learning

Table 1.3: Initial model (online learning)

Parameter	No cues	With cues
Input size	8	10
Number of units	1000	1000
Input connectivity	0.2	0.2
Reservoir connectivity	0.1	0.1
Input scaling	1	1
Spectral Radius	0.8	0.99
Leak Rate	0.1	0.1

2 Chapter 2: Structure is needed for more complex tasks

In the previous chapter, we demonstrated that the minimal model is sufficient to solve a continuous alternating task. We showed that structure is not required for solving this task and that a random RNN is sufficient not only to solve the task but also to exhibit neural properties known to play an important role in spatial navigation. In this section[1], we investigate whether this minimal model can solve other decision-making tasks. To this end, we introduce a new task with varying time constraints, making it more challenging. We will explore how structural variations influence performance on this task, using ESNs as minimal models with varying structures. From a biological perspective, decision-making processes are intimately connected to the basal ganglia network in the brain, which we will first introduce to provide context for this research.

2.1 The Basal Ganglia (BG): the biological decision-making network

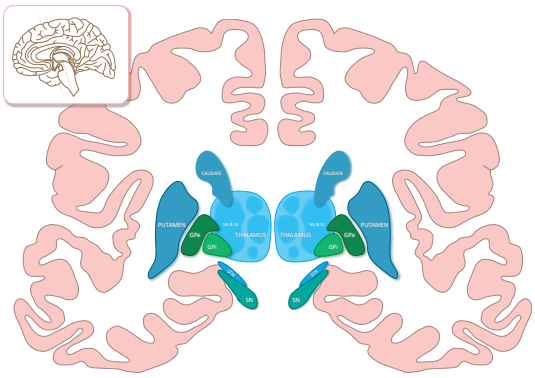

Figure 2.1: The BG are located in the subcortical regions of the brain.

The basal ganglia (BG) are a group of subcortical structures in the vertebrate brain that play a major role in the decision-making process (see Figure 2.1). Their structure have maintained a highly conserved organization for over 500 million years. As illustrated in Figure 2.2, the BG form a closed-loop system with the

[1]This chapter is an extract of Chaix-Eichel et al. [2024].

cortex and thalamus, known as the basal-cortico-thalamic loop, which has been conserved since lower vertebrates.

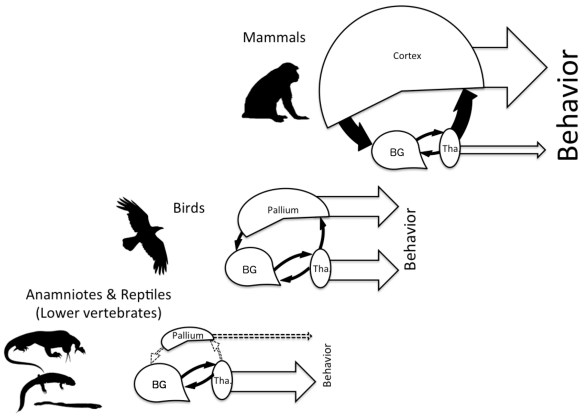

Figure 2.2: Connectivity of the BG with the cortex/pallium and the thalamus, and their roles in behavior expression throughout vertebrate evolution. (From Boraud [2015])

In this chapter, we will begin by providing an overview of the BG anatomy, followed by a detailed explanation of its functional organization. Subsequently, we will delve into the role of dopamine in learning and its interactions with the BG network.

2.1.1 Structure and functions of the BG

The BG are a group of interconnected structures consisting of the striatum (Str), the subthalamic nucleus (STN), the globus pallidus external and internal (GPe and GPi), and the substantia nigra pars compacta and pars reticulata (SNc and SNr) [Mink, 1996], as illustrated in Figure 2.3. The Str and the STN correspond to the input structures of the BG, with the GPi serving as the output structure and the GPe acting as an intermediary structure [Boraud, 2015]. The BG comprise three distinct pathways: the direct, indirect, and hyperdirect pathways, as depicted in Figure 2.3. The direct pathway involves an inhibitory connection from the striatum to the GPi, which transmits inhibitory signals to the thalamus, thereby facilitating movement. The indirect pathway connects the Str to the GPe, from the GPe to the STN, and finally from the STN to the GPi, which then sends inhibitory signals to the thalamus, suppressing movement. Albin et al. [1989] were the first to propose this connectivity model of the BG, highlighting the roles of the direct pathway (Striatum-GPi-SNr) in facilitating movement and the indirect pathway (Str-GPe-GPi/SNr) in suppressing movement. Nambu et al. [2002] extended this model by introducing the hyperdirect pathway, which is crucial for inhibiting competing movements to allow for the selection of the desired action. The hyperdirect pathway forms an indirect connection from the cortex to the thalamus via the STN and then the GPi,

functioning as a fast pathway that inhibits competing movements, thereby enabling the selection of the desired movement. The three pathways modulate the activity of the thalamus, which in turn influences cortical activity, then behavior. The normal functioning of the BG relies on a dynamical balance between the three pathways [Leblois et al., 2006]. This equilibrium is crucial for proper motor control, and any dysregulation can lead to various disorders. For example, Parkinson's disease is characterized by an imbalance between the indirect and direct pathways, which may result from the loss of dopaminergic neurons in the SNc.

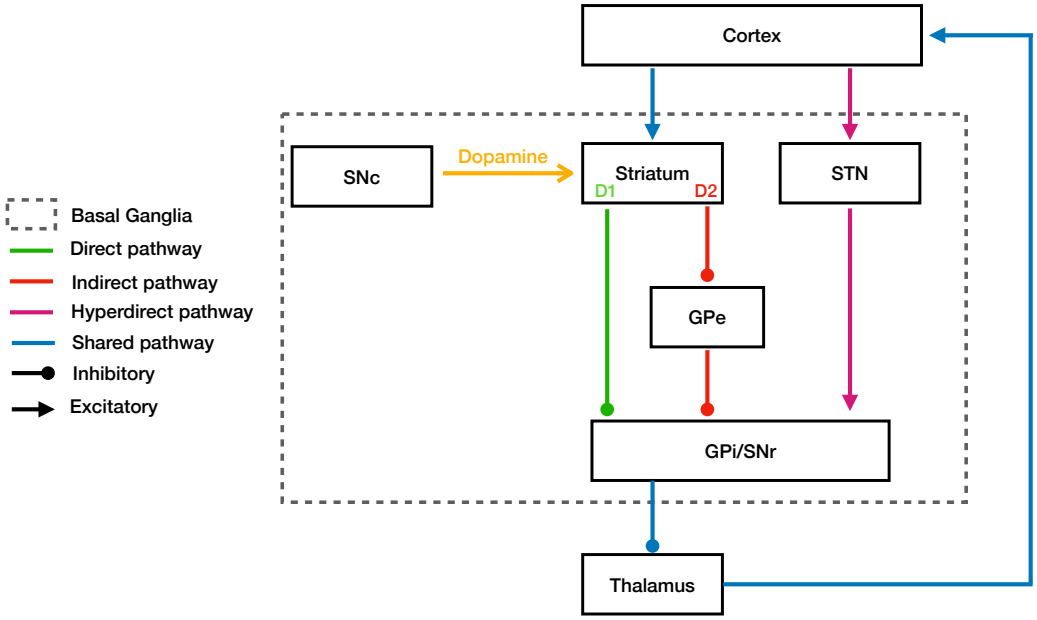

Figure 2.3: BG circuitry. Main pathways through the basal ganglia.

2.1.2 The dopamine system of the BG

Dopamine plays a crucial role in modulating the BG pathways via D1 and D2 dopamine receptors located in the Str [Alexander et al., 1986]. The D1 receptors are primarily associated with the direct pathway, where activation by dopamine facilitates movement by enhancing the inhibitory effect on the GPi. In contrast, the D2 receptors are linked to the indirect pathway, and when activated by dopamine, inhibit the indirect pathway, thereby reducing the inhibitory effect on the thalamus and promoting movement. Understanding the dopamine system in the BG is essential for gaining insights into various neurological disorders. Dysregulation of dopamine can lead to Parkinson's disease, which is characterized by the degeneration of dopaminergic neurons in the SNc. This degeneration results in motor symptoms including tremors, rigidity, bradykinesia, and postural instability [Dauer and Przedborski, 2003]. Additionally, disorders like Huntington's Disease are

associated with dysfunction in dopamine receptors [Ross and Tabrizi, 2011], while Tourette Syndrome is linked to dysregulation of dopamine levels [Kurlan, 1989].

Furthermore, dopamine plays a crucial role in the BG's ability to learn and execute behaviors through reinforcement learning (RL) mechanisms [Schultz et al., 1997]. This neurotransmitter modulates synaptic plasticity within the BG circuits, enabling the system to adapt and refine its action selection processes based on reward prediction errors (RPE), corresponding to the difference between the expected reward and the actual reward. It has been shown that dopamine released by neurons in the SNc and ventral tegmental area (VTA), encodes the RPE [Bayer and Glimcher, 2005]. When there is a positive RPE (the outcome is better than expected), an increase in dopamine release strengthens the synaptic connections involved in the associated choice. Conversely, when the RPE is negative (the outcome is worse than expected), dopamine release decreases, weakening the synaptic connections involved in the selected choice. RPE, and consequently dopamine, play a crucial role in learning and decision-making processes by updating the agent's expectations and behaviors based on the outcomes of its actions.

2.1.3 Preservation of the BG Structure Across Evolution

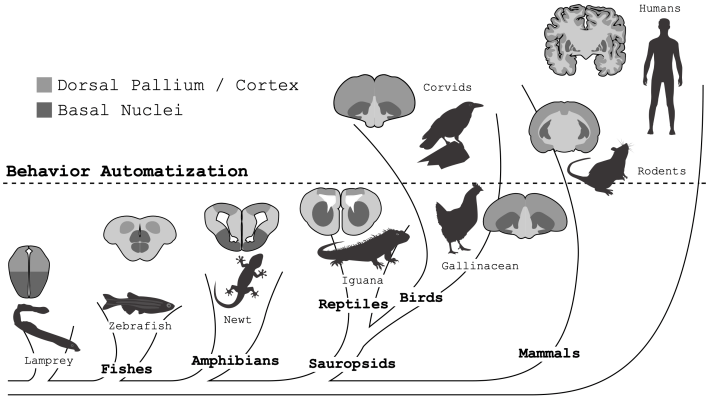

Figure 2.4: Conservation of the BG throughout 500 million years of evolution. (From Boraud et al. [2018])

Studies examining the BG of the lamprey, which diverged from mammals over 500 million years ago, reveal a remarkably detailed and similar organization in terms of connectivity and overall structure [Suryanarayana et al., 2022, Lamanna et al., 2022]. The lamprey's BG contains all the major components found in mammals, including the Str, GP, SNr, and STN [Stephenson-Jones et al., 2011]. These nuclei are interconnected via direct and indirect pathways. The direct pathway involves striatal projection neurons expressing dopamine D1 receptors, while the indirect pathway involves those expressing dopamine D2 receptors [Grillner and Robertson, 2016].

Furthermore, the dopaminergic system associated with reward processing and reinforcement learning in lampreys exhibits striking parallels to those found in mammals. The intrinsic organization, inputs and projections, neurotransmitters, and synaptic membrane properties in the lamprey BG are also similar to those in mammals [Pérez-Fernández et al., 2021]. In essence, the lamprey exhibits a foundational blueprint of the decision-making network [Grillner and Robertson, 2016]. These findings underscore the evolutionary conservation of the BG's structure and function, highlighting its fundamental role in vertebrate behavior and decision-making processes.

2.1.4 Computational Models of the Basal Ganglia

The biological decision-making network refers to the system including the BG, the prefrontal cortex and the thalamus, as depicted in Figure 2.3. These components work together to integrate a wide range of sensory information, processing it to select the most appropriate actions. Over the past decades, this network has been a major focus of computational modeling [Chakravarthy and Moustafa, 2018].

Feedforward models of the BG

Numerous models have been developed using feed-forward architecture as depicted in Figure 2.5 [Humphries et al., 2006, Gurney et al., 2001, Frank, 2006, Bogacz and Gurney, 2007, Collins and Frank, 2014]. These models are composed of the direct (Go) pathway facilitates the selection of the most appropriate action, while the indirect (NoGo) pathway suppresses competing actions, inhibiting other competing motor responses. Earlier models primarily focused on these Go/NoGo pathways, but more recent findings highlight the importance of the hyperdirect pathway, which includes the subthalamic nucleus (STN), [Frank, 2006]. This pathway provides rapid inhibitory signals for precise control over motor responses.

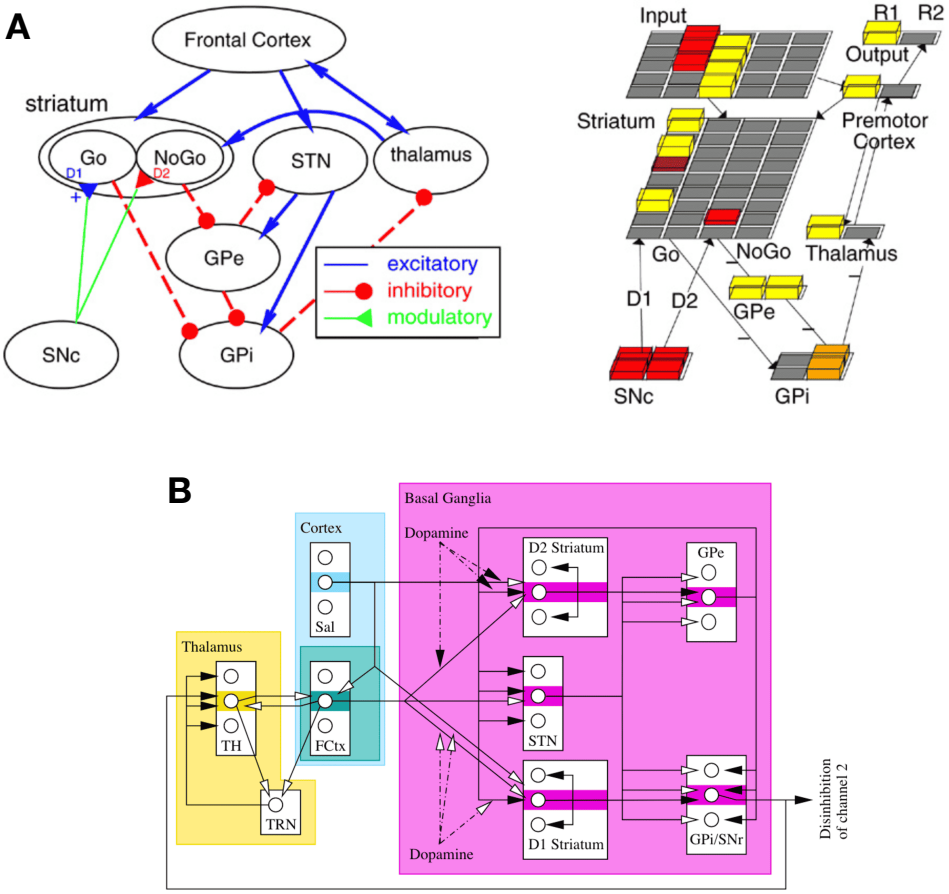

Figure 2.5: Feedforward models of the BG system implying the direct (Go) and indirect (NoGo) pathways. **A**: From Frank [2006]. **C**:From Girard et al. [2008]

BG models based on competing feedback loops: direct and hyperdirect pathways

Other models emphasizes the competition between feedback loops, specifically the direct and hyperdirect pathways, as the primary mechanism for action selection and motor control [Leblois et al., 2006, Guthrie et al., 2013]. The interaction between these pathways is crucial for motor program selection and explains both normal and pathological BG dynamics. In this framework, the Str, STN GPi, thalamus, and cortex form closed feedback loops through the direct and hyperdirect pathways. This is in contrast to previous models that may focus more on feed-forward parallel processing or single-loop dynamics.

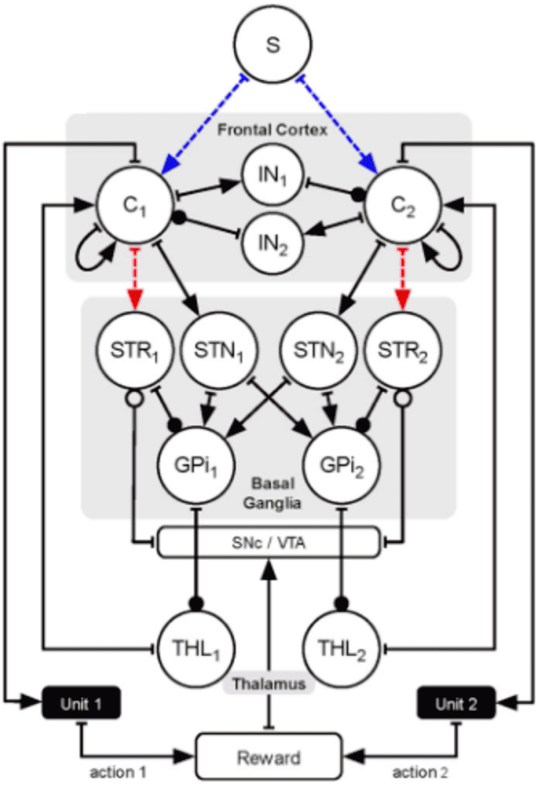

Figure 2.6: Competing feedback loops: direct and hyperdirect pathways. (FromBoraud et al. [2018])

Reinforcement Learning in BG computational models

Dopamine is a key component of the BG system, regulating the action selection process, predicting rewards, and reinforcing behaviors [Maia and Frank, 2011]. RL frameworks [Sutton and Barto, 1998] are common to model cognitive processes [Collins, 2019], and BG functions, particularly in decision-making processes [Collins and Shenhav, 2022, Khamassi et al., 2005]. In machine learning, RL refers to the process in which an agent learns optimal decision-making through direct interaction with its environment as illustrated in Figure 2.7. The agent takes actions, receives feedback from the environment in the form of rewards. In this context, the agent's goal is to find the optimal set of actions (known as policy) that maximizes long-term rewards [Sutton and Barto, 1998]. Two well-known RL models are Q-learning and SARSA [Watkins and Dayan, 1992]. These models are classified as "value-based RL algorithms," which typically estimate the value of each state-action pair using an action-value function (Q-function) or a state-value function (V-function). These

functions indicate how beneficial it is to be in a particular state and take a specific action from that state. Value-based RL algorithms are known as "critic models" because they evaluate the quality of actions taken by the agent. The value is updated based on the temporal difference (TD) error, which measures the difference between the estimated value of the model and the actual reward received:

$$\delta_t = r_t - Q(s_t, a_t) \tag{2.1}$$

where:

- δ_t is the TD error,
- r_t is the reward received at time t,
- $Q(s_t, a_t)$ is the value function when the agent is in state s_t and takes action a_t.

The value is then updated as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \eta.\delta_t \tag{2.2}$$

where η is the learning rate, which controls how quickly the Q-values are updated based on the TD errors. If the TD error (δ_t) is negative, it indicates that the expected value was higher than the actual reward received, leading to a decrease in the value. Conversely, if the TD error is positive, it indicates that the expected value was lower than the actual reward, and thus the value is increased. The dopamine is encoding this TD error (in the simplest case when the reward is immediate), also called RPE. Some models suggest that the value functions is computed and represented in the Str [Samejima et al., 2005, Balleine et al., 2007, Roesch et al., 2009, Ito and Doya, 2015], consequently considered as the 'critic module' of the BG.

However, this view has been challenged. Elber-Dorozko and Loewenstein [2018] argue that the identification of action-value representation in the Str may be erroneous. Moreover, Bennett et al. [2021] propose another branch of RL that directly learns the behavioral policy, without the intermediate step of value-learning, and review neural findings that could be better explained by policy-gradient models rather than traditional value-based models [FitzGerald et al., 2014, Li and Daw, 2011].

Another popular type of RL model is the actor-critic [Konda and Tsitsiklis, 1999] depicted in Figure 2.8, and used to model the BG function [Khamassi et al.]. This model is composed of a critic network, and a second network called the actor. The actor-critic combines policy-based and value-based method: the critic receives the reward, evaluates the value of each states and actions, and sends the learning signal to the actor. The actor uses this evaluation to select the best action. Both networks

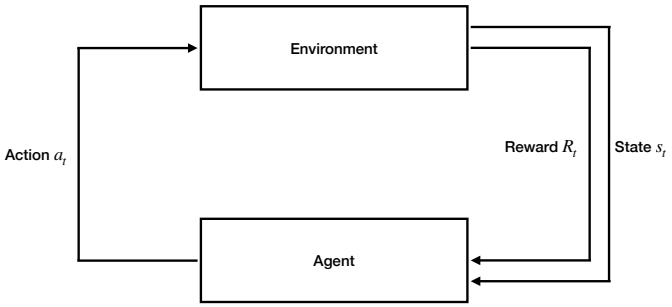

Figure 2.7: The agent-environment interaction in reinforcement learning. The agent takes actions, receives feedback from the environment in the form of rewards.

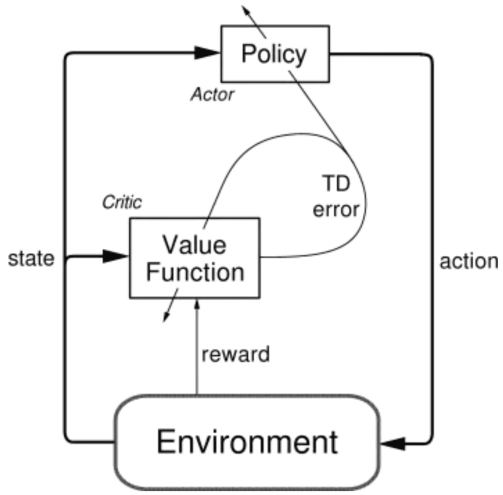

Figure 2.8: Actor-critic architecture. From Sutton and Barto [2018]

receive input from the environment. Several models propose that the basal BG are structurally and functionally analogous to an actor-critic RL model, where different modules within the Str represent the actor and critic components [Houk et al., 1994, Suri et al., 2001, Contreras-Vidal and Schultz, 1999]. However, Joel et al. [2002] discussed these limitations, highlighting the complexity of BG functions and the need for more nuanced models.

In essence, given the well-known and highly conserved structure of the basal BG, often described in terms of pathways, numerous computational models have been built upon its functional and structural anatomy, with the learning process widely acknowledged to be governed by RL. These models provide valuable insights into the mechanisms underlying action selection processes, although the role of each pathway remains under scrutiny. In the next section, we introduce an alternative approach

that diverges from existing methods in the following way: instead of drawing inspiration from the BG structure, we will construct minimal models with randomly connected neurons. This allows for architectures that do not rely on any strong prior assumptions. This unconventional approach aims to explore the broader implications of neural structure on decision-making processes.

2.2 Alternative Decision-Making Network Architectures

In this section, we are primarily interested in exploring the role of structure when solving a decision task while avoiding to make any strong assumption regarding the actual structure. To do so, we exploit the echo state network (ESN) paradigm (see Boxes 0 .1 and 0 .2.) that allows to solve complex tasks based on a random architecture [Jaeger, 2007]. Considering a temporal decision task, the question is whether a specific structure allows for better performance and if so, whether this structure shares some similarity with the BG. We begin with a simple, minimal random model and incrementally introduce complexity to the structure, if doing so proves necessary to improve task performance. Unfortunately, we cannot explore each and every variant of architecture because the number of different structures for a fixed number n of neurons is huge (and grows exponentially with n). Instead, we restrict our exploration to a much smaller subset where a model is built around two pathways, each of them being made of several chained ESN and in charge of processing a single input. We also added a continuous case based on topological reservoir that allows to have distance based connectivity patterns and allows us to take the limit of the two pathways structure. These models are loosely inspired from the direct and hyperdirect pathways of the BG [Schmidt et al., 2013], with the latter allowing the production of a fast "stop signal" thanks to a reactive inhibition. This work has been the subject of a publication [Chaix-Eichel et al., 2024].

2.2.1 The task: time constrained decision-making

Our task is inspired by an experiment involving monkeys coming from Pasquereau et al. [2007], depicted in Figure 2.9. It is a variation of the n-armed bandit task [Auer et al., 2002], where an agent must decide which arm of n slot machines to play in multiple trials to maximize cumulative reward. Specifically, Topalidou et al. [2018] adapted this task for $n = 2$. In the context shown in Figure 2.9, the monkey selects one cue from two options. Subsequently, it receives a reward based on the probability associated with the chosen cue. Additionally, the task clearly distinguishes between the selection of a cue and its associated position. In other words, the final decision is based on the position associated with the cue, rather than the cue identity itself, and this position can change from one trial to another (e.g., in trial 1, the triangle is on the left, but in trial 5, it is on top). This binding

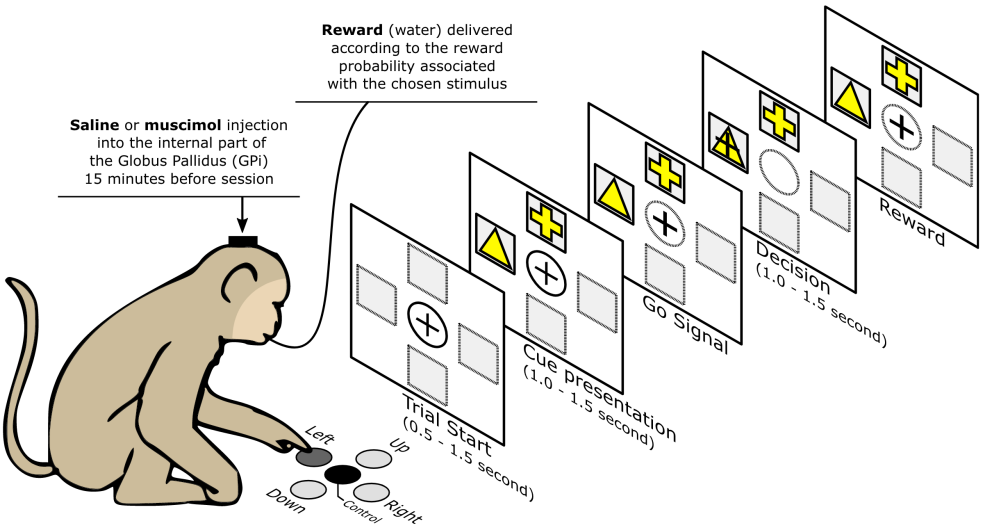

Figure 2.9: Our task takes inspiration of the following behavioral task introduced by Pasquereau et al. [2007]. The monkeys initiate a trial by keeping their hands on the central button. After a random delay, two cues appears in two different positions. The monkey has a random duration time window to press the button associated with one cue. It moves the cursor over the chosen cue and has to maintain the position for some duration. After this delay, the monkey is rewarded or not according to the reward probability of the chosen cue.

problem adds complexity to the task .

We consider a similar task to that of Topalidou et al. [2018], with some modifications to suit our study (see Figure 2.10). We retain the same 2-arm bandit task incorporating the binding problem. However, in our version, the task is non-stochastic: an agent is presented with two options, each associated with a fixed amount of reward.

Motor aspect

As in Topalidou et al. [2018], we maintain the binding problem, meaning the agent's final decision corresponds to the position associated with the chosen stimulus. There are four possible stimuli, each with a distinct identity (triangle, square, circle, losange represented by numbers 1 to 4), and four possible positions (top, down, right, left, also represented by numbers 1 to 4) as shown in Figure 2.10. For any given trial, stimulus identity and position are mutually exclusive. The value of an option is solely attached to the identity of the stimulus, irrespective of its position. The agent's choice is interpreted as selecting a position, from which the identity of the stimulus can be determined (and hence the amount of reward).

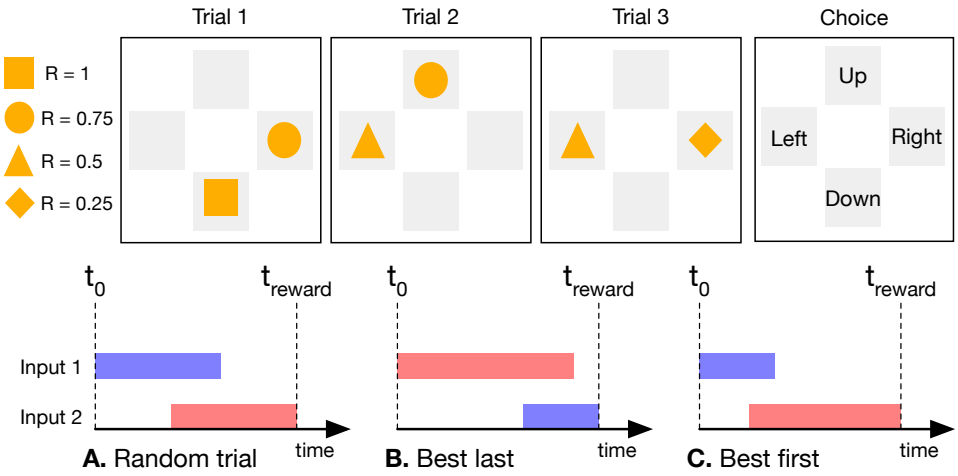

Figure 2.10: **Top**: Three task trials featuring a 4-item cue set with associated reward values. **Bottom**: Stimuli timings. In each trial, the two options appear and disappear with independently random timings. In our case, their appearances overlap, even though they do not start and end at the same time. The red (worst) and blue (best) stimuli can have different onset/offset times, and the reward is received at a fixed time.

Temporal aspect

We consider a time-constrained decision-making task. Each trial has a fixed duration of 30 timesteps, after which the agent receives the reward corresponding to its choice. The reward values are 0.25, 0.5, 0.75, 1, as shown in Figure 2.10. However, within a fixed trial, the onset and offset times of the two options are independent. This means the options may have different durations, may start or finish at different times, and can be completely disjoint, with no overlap between them, as illustrated in the bottom part of Figure 2.10. This temporal aspect significantly increases the task's complexity. In some trials, the agent must retain the value of the first option in working memory, while in others, the agent must contend with a late but better option, navigating the challenges of time-constrained decision making. In this study, we only consider overlapping stimuli.

In the real world, decision making is time constrained [Ordonez and Benson III, 1997]. Decisions need to be taken within certain timeframes, where the importance of speed and the need for caution can vary across situations. In many such cases, there would exist a speed-accuracy tradeoff [Heitz, 2014], where one can collect more information or ponder more over the choice in order to make a better decision at the cost of time. As navigating such trade-offs optimally would be important for

one's survival [Davidson and El Hady, 2019], many sophisticated models have been developed to model animal behavior in such situations. A popular set of models used to study how animals approach time-constrained decisions are Evidence Accumulator Models (EAMs) [Ratcliff and McKoon, 2008, Brown and Heathcote, 2008]. In such models, deliberating over a decision is modeled as accumulating observations which over time can be perceived as evidence for making one or another choice. When the accumulated evidence reaches a certain threshold, a decision is taken. Such models are able to seamlessly integrate the myriads of factors that affect animal decisions, with the threshold indicating response caution, and observations serving as probabilistic likelihoods for choices. Electrophysiological evidence for correlated ramping signals in specific brain regions has been observed [Pisauro et al., 2017]. Another widely used EAMs are the Drift Diffusion Model, which is equivalent to the Wald Sequential Probability Ratio Test. Despite the many factors in favor of EAMs, one occasion where such models fail is when new evidence emerges, necessitating quick decision changes [Cisek et al., 2009]. EAMs struggle in promptly adjusting to sudden changes, especially when the new evidence contradicts previously acquired information. While alternative models such as Leaky Accumulator Models [Usher and McClelland, 2001] and Urgency Gating Models [Thura et al., 2012] have been proposed as a solution to this problem, they often don't provide as good fits to animal behaviour when considered across a wide variety of tasks. This paper introduces an alternative approach to address this challenge, with the objective of building ESNs with multiple architectures, each capable of handling temporal information in a distinct manner.

2.2.2 Derivation of ESN random architecture

Different models trained with online RL

We aim to explore various architectures derived from ESNs. Specifically, we modify the random reservoir component and replace it with alternative structures to assess the impact on a specific decision task. The role of structure in ESN has already been addressed in a number of works. While Dale et al. [2021] quantified how structure affects the behavioral characteristics of the ESN, several studies have demonstrated that replacing the initial random topology of the ESN by more organized structures could improve the overall performances of the model. Nonetheless, rather than completely removing the randomness of the network topology, certain structures allow to combine both random and structured connections. One well-known example is the small world network, which has been observed in the neural network of the C. Elegans [Watts and Strogatz, 1998] and in other brain systems [Bassett and Bullmore, 2006]. Cheng et al. [2015], Bai et al. [2017] have shown that incorporating small-world structure into ESNs results in performance improvements on benchmark tasks. Various other structures for ESN have also demonstrated significantly superior performance, including the combination of scale-free and small-world networks [Deng and Zhang, 2007, Kawai et al., 2019, Kitayama,

2022]. Additionally, modular structures [Rodriguez et al., 2019], forward topology with shortcuts pathways [Dominey et al., 2022] and hierarchically clustered ESNs [Jarvis et al., 2010] have been explored, each impacting memory capacity, temporal properties, and reservoir stability. Biological neuronal networks (BNNs) have been utilized to implement reservoir computing with the neurons in cultured neuronal networks randomly connected [Sumi et al., 2023], but with a modular topology [Takuma et al., 2023], demonstrating increased reservoir computing performance. An alternative approach known as Deep Reservoir Computing, involves investigating various structures through the combination of multiple random reservoirs rather than a single one [Gallicchio et al., 2017, Moon et al., 2021, Xue et al., 2007, Manneschi et al., 2021a, Stenning et al., 2023]. Our study aims to contribute to the existing findings by incorporating the application of Deep Reservoir Computing and the utilization of ESN with a forward topology.

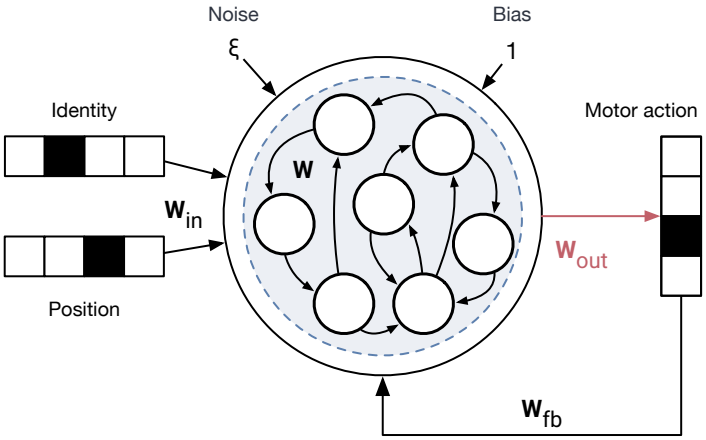

Figure 2.11: **Top:** Model architecture. An ESN with a motor output (movement direction from 1 to 4). The model comprises a pool of randomly interconnected neurons in the reservoir. Black arrows indicate fixed weights, while the red arrow indicates plastic weights, trained with RL. There is also a feedback connection from the output to the reservoir.

We start with this ESN model with an initial random architecture, as depicted in Figure 2.11. The model takes as input the identity and position of the two stimuli and has to output its motor action, which corresponds to the position related to the chosen stimuli. The readout layer is trained using online RL based on equation 3.1 and 3.2, where only the weights associated with the selected choice undergo updates. The choice of RL as the learning rule comes from its biological plausibility, given that cortico-basal-ganglia (BG) circuits are trained through reinforcement, thanks to the encoding of reward prediction error (RPE) with dopamine [Bar-Gad et al., 2003]. Equations read:

$$W_{out}(\text{choice}) = W_{out}(\text{choice}) + \delta W_{out} \tag{2.3}$$

$$\delta W_{out} = \eta \cdot (r - \text{softmax}(\mathbf{y}, \beta)[\text{choice}]) \cdot (\mathbf{x} - \mathbf{x}_{\text{th}}) \tag{2.4}$$

where *choice* represents the index associated with the model's chosen action. η is the learning rate. The function $softmax(y, \beta)$ applies softmax to the model's output with y as the model's output and β as a parameter. x_{th} denotes a small constant value, and r corresponds to reward feedback received. Each stimulus identity is associated with a specific fixed reward value, which can be 1, 0.75, 0.5, or 0.25. The learning rule is inspired from the work of Zhang et al. [2018]. The action selection process follows the epsilon-greedy method, allowing to balance between exploitation and exploration phases. This balance is crucial for optimal learning, as highlighted by Velentzas et al. [2017], Khamassi et al. [2010] when learning a task . When the agent is in the exploitation mode, it selects the action that corresponds to the highest output value of the model ($argmax(output_{model})$). In contrast, during exploration, the agent randomly selects one action from the set of all available actions, with equal probability among the four possible choices.

The method uses a parameter called epsilon (ϵ), which starts at 1 during the beginning of each simulation and ends at 0, signaling a shift towards exclusive exploitation of learned knowledge. This dynamic ϵ adjustment enables the agent to transition from exploration to exploitation.

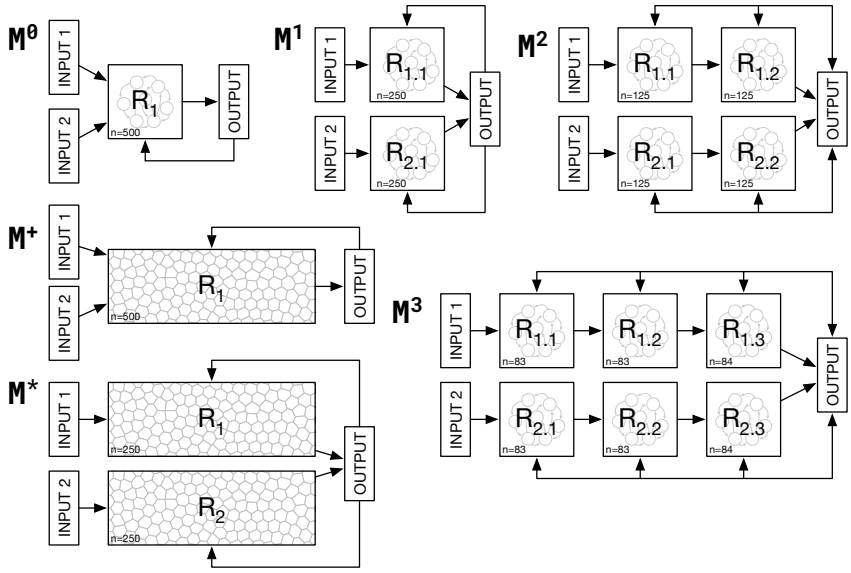

Figure 2.12: Model are composed of one to several chained ESN, all connected to the readout, all receiving feedback from the output. $\mathbf{M^0}$: Regular ESN. $\mathbf{M^1}$: Dual pathway each made of a single ESN. $\mathbf{M^2}$: Dual pathway each made of two chained ESNs. $\mathbf{M^3}$: Dual pathway each made of three chained ESNs. $\mathbf{M}^+$: Continuous ESN. $\mathbf{M}^*$: Dual pathway each made of one continuous ESN.

From this classical ESN (Figure 2.11), we derived several architectures (Figure 2.12) that are all characterized by the presence of two distinct pathways, a slow

pathway and a fast pathway, drawing inspiration from the direct and hyperdirect pathways of the basal ganglia [Schmidt et al., 2013], with the latter allowing the production of a fast "stop signal" thanks to a reactive inhibition. This "stop emergency brake"[Aron and Poldrack, 2006] is attributed to the significant role of the nucleus STN. Each of these two pathways receive segregated inputs, that is, each pathway receives a single option. More precisely, the slow pathway receives the earliest option and the fast pathway receives the latest option. All the models possess a total of 500 neurons equally distributed across multiple reservoirs, with all reservoirs connected to the readout and receive feedback signals from it.

We also designed two models (M^+ and M^*) that are equipped with a topology [Rougier, 2018] such that it is possible to constrain activity propagation along a feed-forward axis (from input to output). This allows the reservoir to progressively process information along the main axis, where early units (that are closer to the input) have access to local and recent information while late units have access to global and processed information. The output layer which has access to both early and late units has the ability to accumulate information and take accurate decisions, while at the same time having the ability to quickly respond to changes in the environment. To make these type of reservoir, the distribution neurons across a 2D space is first defined by using the algorithm described in [Rougier, 2018] from which the connectivity matrix can be derived. Individual connections are established based on local connectivity rules [Dominey, 2024, Ceni et al., 2024], where the nearest neurons meet angle constraints as shown in figure 2.13, connections are established between input and output neurons following a rule in which the probability of connection exponentially decreases with distance.

Model optimization

All models undergo a hyperparameter optimization process using the Optuna library [Akiba et al., 2019]. More specifically, spectral radius (sr), leak rate (α), input connectivity (W_{in}), output connectivity of W_{out}, the reservoir connectivity of W, exploration rate (ϵ), the temperature of the softmax (β) and the learning rate (η) are optimized. For the $\mathbf{M}^*$ model, rather than the spectral radius and reservoir connectivity, the connections are determined by the radius, angle and sparsity parameters which are optimized instead. The spectral radius parameters is only applicable when the connection angles are set to a value greater than 90°, as when they are lesser than 90°, there are no recurrent connections in the reservoir, and thus no possibility of chaotic activity as any input would inevitably decay. In the case of angles greater than 90°, we found that constraining the weights based on spectral radius had a detrimental effect on the performance of the network. We believe that due to the input being presented from one side of the network and the unique nature of connectivity in the reservoir, not all eigenvectors of the reservoir connectivity matrix may become instantiated in the network. Thus, as scaling the weights based on the largest eigenvalues were not giving the best results, the weights

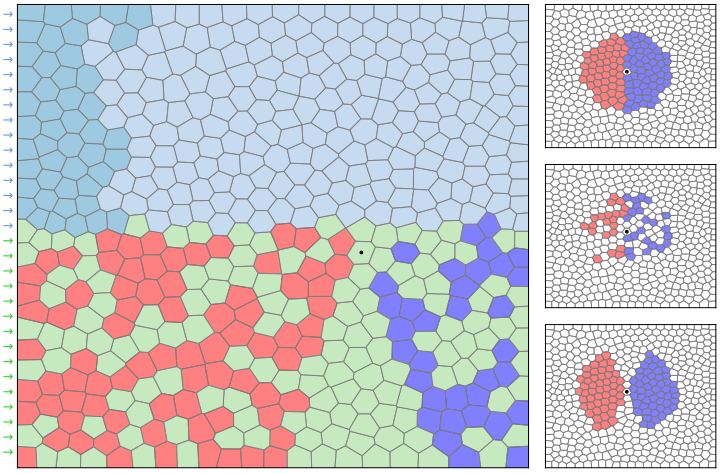

Figure 2.13: **Left** Internal structure of the M* model. The two pathways are shaded in blue (top) and green (bottom) and are completely segregated (no reciprocal connections between them). The input to the top pathway has ben shaded in darker blue while the output is not represented for clarity because output receives connection from virtually all units. The input to the bottom pathway is similar but not represented for clarity. Instead, a typical unit (dot) connection pattern is represented with red for incoming connections and blue for outgoing projections. **Right** The connection pattern of a unit is governed by an angle θ (ranging from 0 to 90), a fixed radius r and a connection probability P_c. Top) $\theta = 90°$, $P_c = 1.0$ Middle) $\theta = 90°$, $P_c = 0.4$ Bottom) $\theta = 70°$, $P_c = 1.0$.

of the topological reservoirs have not been scaled using a spectral radius in this work. The optimization process consists in running 600 simulations with different set of parameters sampled using Tree-structured Parzen Estimator (TPE) [Bergstra et al., 2011]. Each simulation consists of 1000 trials, and performance assessment occurs over the last 200 trials of the simulation by counting the number of successful actions (best option chosen).

Protocol

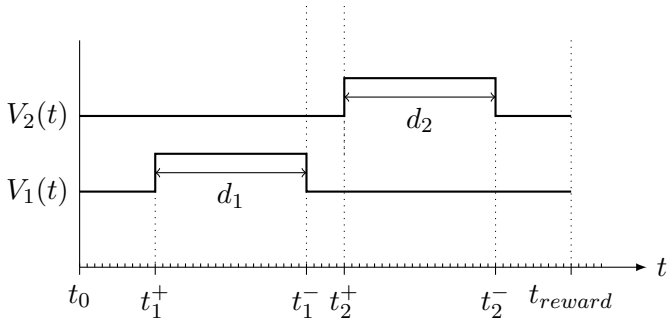

Figure 2.14: The two stimuli V_i are characterized by their respective onset (t_i^+) and offset (t_i^-) time. The time of decision t_{reward} is fixed and constant across trials.

The models were optimized across a broad spectrum of timing and delay conditions (see Figure 2.14). While $t_{t_1^+}$ and t_{reward} are fixed and respectively set to 5 and 1, d_1 and d_2 vary within the range of 5 to 20, and $t_1^+ - t_2^+$ fluctuates between 0 and 20, with these values being randomly generated from trial to trial. This approach enables to identify the optimal parameters that yield superior performance across all potential delay scenarios. This performance assessment is designed to quantify the extent to which the models demonstrate temporal task generalization capabilities.

Analysis

After optimization, the models are trained to select the position associated with best stimuli (equivalent to the most rewarding one). The training procedure consists of 1000 trials, with each trial being randomly chosen from the 72 stimulus-position pairs. The timings and delays for each stimulus are also randomly determined. Following the training phase, the models undergo testing on 1000 randomly selected trials. The overall performance is measured as the proportion of correct choices out of the 1000 trials. The results are further analyzed by separating two scenarios: when the best stimulus appears first and when the best stimulus appears last. The first scenario enables an evaluation of the models' working memory: if the best stimulus emerges first, the model must retain its value until the end of the trial. The second scenario allows an evaluation of performance when the model needs

to respond rapidly: if the best cue appears last, the model must quickly adjust its decision before the trial ends. This entire process is repeated across 10 different seeds for each model, and the final performance is determined as the average across the seeds. A paired t-test is employed to assess whether there is significant differences between the performances of the model M^0 and the other ones.

Results

The results depicted in Figure 2.16 indicate that the models M^2, M^3, and M^* exhibit significantly better overall performances compared to models utilizing a single reservoir such as M^0 (blue bars). The paired t-test with M^0 results in p-values of 3e-4, 3e-5 and 1e-8 respectively. These three models demonstrate overall better performances primarily because they outperform in the scenario where the best cue appears first (green bars), whereas no particular difference is observed in the scenario when the best cue appears last (yellow bar). These models have the common thread of being composed of two pathways with distinct average leak rates that emerged from the hyperparameter optimization of the models, and depicted in Table 2.1. The difference in values of leak rates between pathways enables different speed of treatment: the latest cue is always fed into the pathway P2 that has a faster processing thanks to a bigger leak rate. Conversely, the earliest cue is always fed into the pathway P1 that has a slower processing thanks to a small leak rate. Furthermore, the overall performances are improving as the depth of the pathways is increasing, until reaching the continuous limit with M^*. The latter achieves the best performances with 89.5% of success, outperforming the classical reservoir with 74.0% of success. This improvement is also mostly visible in trials when the best cue appears first, going from 68.2% of success with M^1 to 87.8% of success with M^*. The difference in performances is visible during the training process as depicted in Figure 2.15, where the dual models are learning faster as the depth of the paths is increasing.

Model	**Pathway 1 (P1)**	**Pathway 2 (P2)**	**P2 / P1**
$\mathbf{M^0}$	0.06	—	—
$\mathbf{M^1}$	0.068	0.67	≈ 9.8
$\mathbf{M^2}$	0.17 (0.06, 0.28)	0.29 (0.50, 0.07)	≈ 3.6
$\mathbf{M^3}$	0.23 (0.16, 0.10 ,0.43)	0.59 (0.07, 0.72, 0.99)	≈ 2.5
$\mathbf{M^*}$	0.23	0.59	≈ 2.5

Table 2.1: Mean value of the leak rates (α of equation 3.1) for pathway 1 & 2 in the models. The ratio $P2/P1$ is highlighting the existence of a fast (characterized by a big leak rate) and a slow pathway (characterized by a low leak rate).

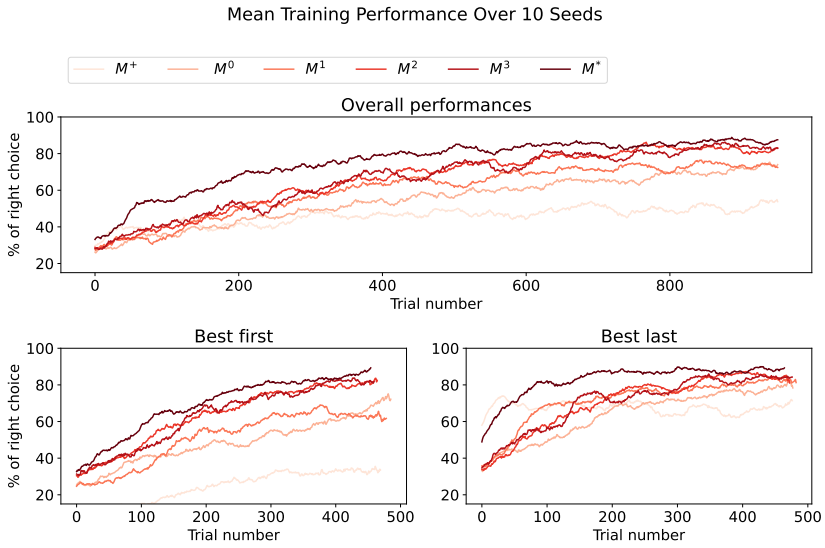

Figure 2.15: **Top** Training process. The curves correspond to the percentage of successful choice using a moving average that takes the 50 last trials. **Bottom** The training process is categorized in two scenarios based on when the best cue appears. This enables to observe that the major difference in performances occurs when the best cue appears first.

2.2.3 Discussion

Starting from a trivial and abstract non-stochastic two-arm bandit task, we complexified the task by introducing a motor indirection as well as a temporal component (embodiment). The initial task (no motor indirection, no temporal aspect) can be fully solved by an ESN in just a few trials using reinforcement learning (with a success rate of 96.7% when motor indirection is absent and 99% when temporal aspect is absent). However, as soon as motor indirection and temporal constraints are jointly introduced, performances dropped drastically, especially in the case when the best option is presented first. These surprising results can be understood when considering a few representatives cases (see also Fig. 2.17):

- When the best option is presented first, it is not sufficient to memorize the motor action to be made because when the second stimuli arrives, the expected value of the first stimulus (as computed internally by the model) needs to be compared to the expected value of the second stimulus such as to make the right motor command.

- When the best options is presented late, the model has only a few time steps to process the new option in order to find the related expected value, to compare this value to the current one and finally, to make a motor decision that needs to overcome the alternative choice.

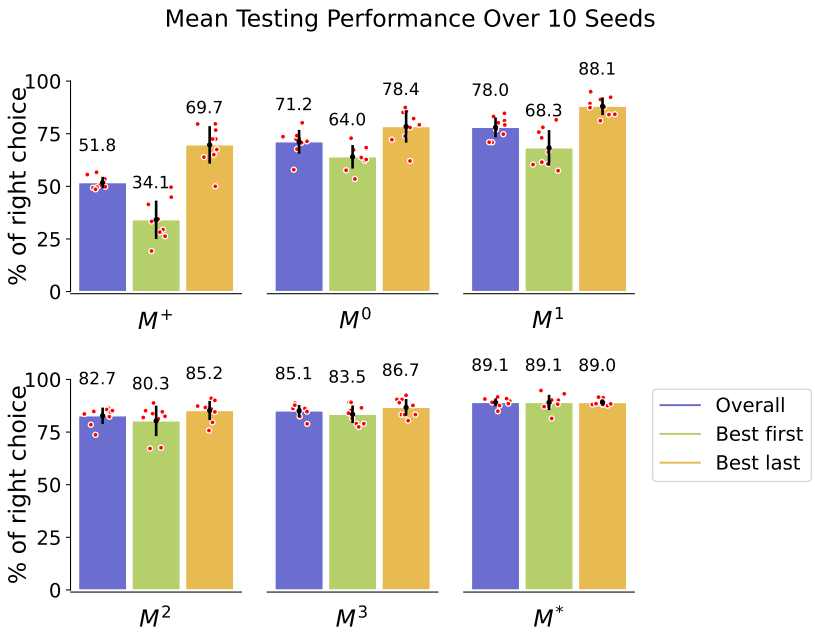

Figure 2.16: Performance comparison of the models. The blue bar corresponds to the percentage of successful choice out of the 1000 tested trials. This result is categorized in two cases: when the best cue appears first (green) and when the best cue appears last (yellow). Models with dual pathway and sufficient depth such as M^2, M^3 and M^* gives significantly better performances overall and when the best cue appears first. M^* corresponds to the deepest dual pathway and demonstrates optimal performances.

Said differently, the model needs to be reactive for some trials and conservative for some others. Results displayed on figure 2.16 clearly indicate that this is hardly the case for the regular $\mathrm{M}^{\mathbf{0}}$ model, with a mean performance of 74%. However, as soon as we introduce a dual pathway architecture, performances increases with the depth of the pathway, best performances being achieved by the continuous M^*.

Table 2.1 displays the mean leak rates over the two pathways in all models (when relevant). Interestingly enough, the second pathway, that is, the one receiving the late stimulus has a much stronger leak rate when compared to the first pathway. This means units in the second pathway are able to process information much more rapidly when compared to the first pathway, even tough this also mean they have a reduced memory capacity (because they leak past activity at a high rate). Overall, all the dual pathways models ($\mathrm{M}^{\mathbf{1}}$, $\mathrm{M}^{\mathbf{2}}$, $\mathrm{M}^{\mathbf{3}}$, M^*) developed both a slow and fast pathway resulting from the optimization process. If we now turn back to our initial inspiration on the structure of the basal ganglia, this dual slow/fast pathway is reminiscent of the direct and hyperdirect pathway even though we do not pretend for equivalence.

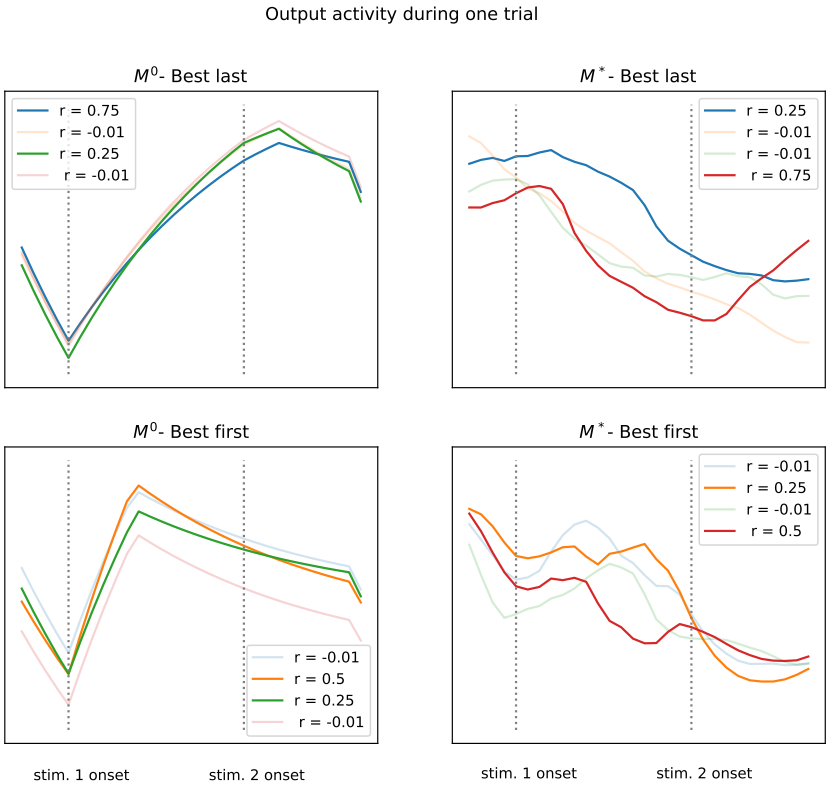

Figure 2.17: Output activity during challenging trials where the reward have close values. **Top** The $\mathbf{M^0}$ fails at processing the best cue that arrives late, while $\mathbf{M}^*$ model is able to react quickly with the late best option and to choose the corresponding best motor action. **Bottom** The $\mathbf{M^0}$ fails at retaining the best cue that arrives first while $\mathbf{M}^*$ model successfully recalls the best option and to choose the corresponding best action.

These models are well suited for tasks involving two options, and emphasize the importance of having multiple timescales of processing. This allows for dealing with different timings constraints, different speeds of response and memory retention. Nevertheless, for tasks like n-arm bandit with n>2, additional pathways are required, i.e. one for each option, with one pathway operating for distinct timescales. This highlights a limitation in generalizability of our models when applied to more complex decision-making scenarios. A key future direction would be to adapt these models to more generalizable decision task with varying time constraints, and apply cross-validation.

3 Chapter 3: Exploring the Necessity of Structure and Multiple Realizability

In the previous two chapters, we demonstrated that an unstructured network is sufficient for solving simple navigation tasks. However, as the complexity of tasks increases, additional structure becomes necessary for making accurate decisions. In this chapter, we explore whether this required structure is unique or if different structures can successfully achieve the same task. In other words, we are exploring whether the concept of multiple realizability - the idea that a particular function can be achieved by different systems [Putnam, 1967] - is at play. When this concept is applied to structures, it is referred to as degeneracy [Edelman and Gally, 2001].

To address this question, we will examine degeneracy in the context of economic decision-making, where choices are made by weighing factors such as costs, benefits, and subjective preferences. In many scenarios, one option clearly stands out as the best choice. For example, if asked to choose between receiving $50 or $100, the choice is straightforward. However, in other situations, the options presented may be objectively equivalent. Consider a scenario where an individual must choose between two options: in the first, they can receive $100 by flipping a coin and landing on tails; in the second, they receive $50 immediately. Both options have identical expected values: 0.5 *100 = 50 in the first case, and 1 * 50 = 50 in the second. In this case, the choice depends on subjective preferences. Some individuals may be risk-seekers and opt for the first option, while others may prefer the certainty of the second option, exhibiting risk-averse behavior.

In similar scenarios, various factors can influence this subjective choice. We aim to explore whether the structure of decision-making networks can influence the behavior in the situations involving risk-benefit tradeoffs. This choice is motivated by the fundamental role that risk attitudes play in shaping behavior across all species [McDermott et al., 2008]. Risk-taking propensities may represent a significant driver of evolutionary processes [Trimpop, 1994, Salas-Rodríguez et al., 2022], making it a crucial area of study. Taking risks, even life-threatening situations, might have been advantageous in certain environment, being essential to survival or reproduction [Steinberg, 2017]. Trimpop [1994], suggests that risk-taking behavior plays a crucial role in evolution and is at least partially influenced by genetics.

With this perspective, our study aims to address the following questions:

- Can different structures result in the same risk-taking behavior, in line with the concept of degeneracy?
- Can the structure itself influence the risk-taking behavior?

To address these questions, we will train a ESN model with different architectures to perform an economic task involving trade-offs between risk and benefits. Given the strong connection between risk-taking behavior and evolutionary processes, we will use genetic algorithms to evolve the model's architecture, enabling the exploration of a wide range of potential solutions for the task. We will then analyze the converging structures that emerge from this evolutionary process and examine their risk-taking behavior in the decision-making task.

3.1 Risk-taking behavior in Economic Decision-Making Tasks

3.1.1 Initial task setup

The task employed in our models ([Nioche et al., 2019, 2021]) was originally designed for monkeys. It is illustrated in Figure 3.1. At every trials, two lotteries $L_{right} = (x_{right}, p_{right})$, and $L_{left} = (x_{left}, p_{left})$ were displayed on a screen. If the monkey chooses L_{right}, it will gain the quantity x_{right} with a probability p_{right}, and will gain the quantity x_{left} with a probability p_{left} if it chooses L_{right}. The monkey has to select the most rewarding one to maximize gains and to minimize losses.

The quantities x_i are represented with line orientation as represented in Figure 3.1: horizontal lines represent 0; clockwise rotation represents a loss of -1, -2, -3 tokens; counter-clockwise rotation represent a gain of 1, 2, 3 tokens. Each lottery is represented by a piechart composed of 2 slices, each slice encode one possible outcome of the lottery (x or 0), and the arc length of each slice represents the probability of the corresponding outcome (p or 1 - p). The possible probabilities are 0.25, 0.50, 0.75, 1.00. The monkey starts each trials with three tokens displayed on the screens. According to the outcome of the trial, the monkey can loose or gain tokens. At the end of the trial, the monkey will receive a liquid reward proportional to the number of tokens earned.

The task consists of two distinct trial groups [Nioche et al., 2019]. The first group, designated as the control group, serves to evaluate the monkeys' ability to perform the task well. This control group comprises five types of trials:

- Type 1: $x_1 > 0$ and $x_2 < 0$ while $p_1 = p_2$, lottery pairs containing one lottery with potential losses and one lottery with potential gains).

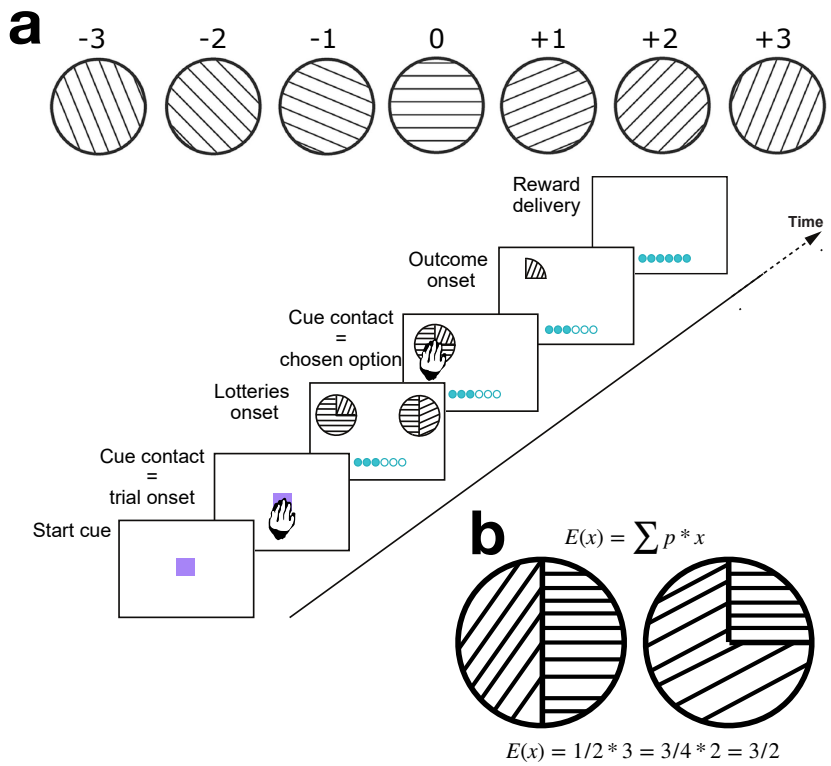

Figure 3.1: **Economic decision-making task** a: The orientation of the lines indicates a quantity. Each lottery is represented by a pie chart composed of two slices. The arc length of each slice represents the probability of the corresponding outcome (p or 1-p). b : Two different pie charts with the same expected value. If the monkey chooses the left pie chart, it will have half chance of getting 3 tokens, if it chooses the right pie chart, it will have 75% chance to get 2 tokens. While left pie chart gives better outcome, it is riskier choice than the right one. A tradeoff has to be made between risk and quantity. (Figures is taken from Nioche et al. [2021].)

- Type 2: $p_1 = p_2$ and $x_1 > x_2$, with $x_{i\in\{1,2\}} > 0$, lottery pairs with potential gains with a stochastic dominant option differentiating only by the x values.
- Type 3: $p_1 = p_2$ and $x_1 < x_2$, with $x_{i\in\{1,2\}} < 0$, lottery pairs with potential loss, with a stochastic dominant option differentiating only by the x values); assess the discrimination of negative quantities; 12 different lottery pairs.
- Type 4: $p_1 > p_2$ and $x_1 = x_2$, with $x_{i\in\{1,2\}} > 0$, lottery pairs with potential gains with a stochastic dominant option differentiating only by the p values).
- Type 5: $p_1 < p_2$ and $x_1 = x_2$, with $x_{i\in\{1,2\}} < 0$, lottery pairs with potential loss with a stochastic dominant option differentiating only by the p values).

The second group of trials is designed to assess the risk-taking behavior of the monkeys. This group includes two types of trials:

- Type 6: $p_1 < p_2$ and $x_1 > x_2$, with $x_{i\in\{1,2\}} > 0$, lottery pairs with potential gains with no stochastic dominant option).
- Type 7: $p_1 < p_2$ and $x_1 < x_2$, with $x_{i\in\{1,2\}} < 0$, lottery pairs with potential loss with no stochastic dominant option).

In the second group, specific situations arise where the expected values of the two lotteries are equal, meaning the two lotteries are theoretically equivalent. An example is illustrated in Figure 3.1-b, where the pie charts for both lotteries show the same expected value. If the left lottery is chosen, there is a 50% chance of receiving 3 tokens, while selecting the right lottery offers a 75% chance of winning 2 tokens. Although the left option potentially yields a higher reward, it involves greater risk compared to the right one. Therefore, a tradeoff must be made between risk and potential payoff. In the next chapter, we will explore how Prospect Theory (PT) provides a powerful framework for assessing behavior in risky situations. This theory reveals consistent behavioral patterns in both humans and other animals, showing that humans tend to be risk-averse when dealing with potential gains, often opting for the safer choice, and risk-seeking when faced with potential losses. These patterns will be thoroughly examined in the following chapter.

3.1.2 Task formalisation for simulations

We formalize this task such as to make it compatible with the ESN models. For simplicity, only the gain scenario is considered where the possible values are 1, 2, 3, and the possible probabilities are 0.25, 0.5, 0.75, 1. The lottery is represented abstractly with rows representing values and columns representing probabilities. Each matrix cell holds the value 1 if the lottery has a probability p (column) and a quantity x (row). The matrices has the following shape:

	$p = 0.25$	$p = 0.5$	$p = 0.75$	$p = 1$
$x = 3$	.	.	.	.
$x = 2$	.	.	.	.
$x = 1$	.	.	.	.

Figure 3.2: Description of the task.

This abstraction avoids explicitly specifying the value and probability for each instance, analogous to the monkey task where monkeys must learn the meaning of line orientations and segments in terms of probability and quantity. For instance, the two lotteries shown in Figure 3.1-b are associated with the following matrices:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 3.3: Left: a matrix representing a lottery with a 50% chance of gaining +3 and a 50% chance of gaining 0. On the right: another matrix representing a lottery with a 75% chance of gaining +2 and a 25% chance of gaining 0.These two matrices correspond to the two lotteries depicted in Fig. 3.1.

In each trial, the model is presented with two matrices as inputs and must learn to select the most rewarding lottery using reinforcement learning (details to follow). The trials are processed sequentially over 8 timesteps: during the first 3 timesteps, no stimuli are presented (null matrices), representing an initialization phase without any signals. This is followed by 5 timesteps during which the two stimuli are displayed. Initially, the model learns to choose the most rewarding lotteries in the control case (control group consisting of lotteries of types 2 and 4, focusing only on the gain scenario). Subsequently, the model's risk-taking behavior is evaluated by introducing lottery type 6, which is specifically designed to assess the model's propensity for risk. After the model has been trained and tested with lottery 6, its risk-taking behavior can be analyzed by examining the proportion of safe versus risky choices it makes in scenarios where the expected values are equivalent.

3.2 Evolving Structures using genetic algorithms

Genetic algorithms have shown to be a powerful tool to optimize hyperparameters of ESNs in time series prediction tasks[Léger et al., 2024]. Considering evolutionary aspects, we propose to extend the application of genetic algorithms beyond hyperparameter optimization to evolve the structure of the ESN model itself. To do so, we use the NeuroEvolution of Augmenting Topologies (NEAT) method [Stanley and Miikkulainen, 2002], which evolves neural network topologies using a genetic algorithm combined with a technique called 'augmented topologies.' This genetic algorithm provides a mechanism to design the topology of a neural network. This approach is applied to ESNs [Chatzidimitriou and Mitkas, 2010, Matzner, 2017].

3.2.1 Model architecture

We use a similar ESN model to the one described in Chapter 2 (Subsection 2.2.2), with a randomly initialized architecture, as illustrated in Figure 3.4. The model takes as input the two lotteries and outputs a choice between them. Based on the reward and probability of the selected lottery, the model receives a reward feedback used to train the readout layer. The later is trained using online RL based on equation 3.1 and 3.2, where only the weights associated with the selected choice

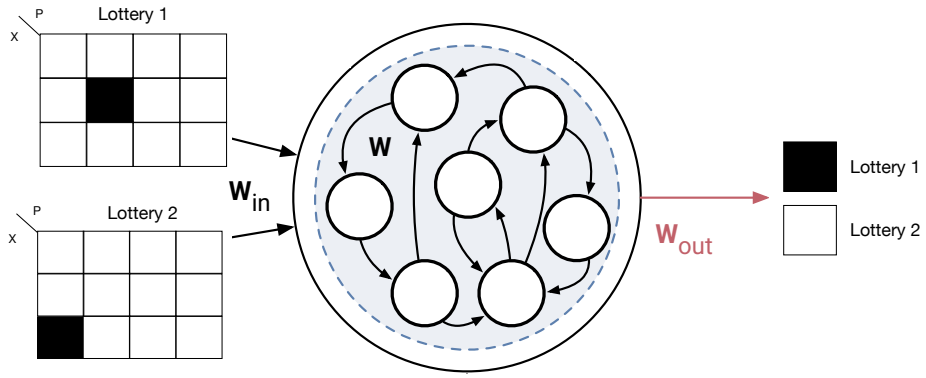

Figure 3.4: Model architecture. Takes as input the two lotteries and outputs whether it selects output 1 or 2.

undergo updates.

$$W_{out}(choice) = W_{out}(choice) + \delta W_{out} \tag{3.1}$$

$$\delta W_{out} = \eta.(r - \mathbf{y}, \beta)[choice]).(\mathbf{x} - \mathbf{x}_{th}) \tag{3.2}$$

where *choice* represents the index associated with the model's chosen action. η is the learning rate, y is the model's output, x_{th} denotes a small constant value, and r corresponds to the reward received. Each stimulus identity is associated with a specific fixed reward value, which can be 1, 0.75, 0.5, or 0.25. The action selection process follows the epsilon-greedy method, allowing to balance between exploitation and exploration phases. When the agent is in the exploitation mode, it selects the action that corresponds to the highest output value of the model ($argmax(output_{model})$). In contrast, during exploration, the agent randomly selects one action from the set of all available actions, with equal probability among the four possible choices. The method uses a parameter called epsilon (ϵ), which starts at 1 during the beginning of each simulation and ends at 0, signaling a shift towards exclusive exploitation of learned knowledge. This dynamic ϵ adjustment enables the agent to transition from exploration to exploitation.

Similarly as in Chapter 2, the model undergoes a hyperparameter optimization process using the Optuna library [Akiba et al., 2019] when running on the control lotteries (lotteries 2 and 4). More specifically, spectral radius (sr), leak rate (α), input connectivity of W_{in}, the reservoir connectivity of W, exploration rate (ϵ) and the learning rate (η) are optimized. The output connectivity is set to 1. Throughout the evolution of the reservoir structure when applying the genetic algorithm, the spectral radius, as well as the input and output connectivity, will change in tandem with the structure, while the other parameters will remain fixed.

3.2.2 Introduction to NEAT

NEAT algorithm was first introduced by [Stanley and Miikkulainen, 2002]. It consists in evolving the topologies dynamically: nodes and connections are incrementally added to the network. In the original case, the ESN begins with a minimal structure (no internal connections) and gradually evolves into more complex neural architectures.

The NEAT algorithm is built on the principles of evolutionary algorithms (EAs) [Eiben and Smith, 2015], more precisely on genetic algorithms (GAs) [Mirjalili and Mirjalili, 2019]. The fundamental process is the following:

- **Initial Population generation**: the process begins with the generation of a random population of individuals, where each individual represents a unique neural network topology.
- **Fitness Evaluation**: each individual in the population is subsequently evaluated by calculating its fitness.The fitness function evaluates how well each genome performs the task thanks to some metrics. In reinforcement learning contexts, it is often equivalent to the cumulative reward or score that the neural network achieves.
- **Selection**: individuals are selected to become parents for the next generation based on their fitness values. Most of the time, this selection process favors individuals with higher fitness, increasing the likelihood that their advantageous characteristics will be passed on.
- **Reproduction**: offspring are produced through genetic operators, which include:

  Crossover: the combination of genetic information from two parent individuals to create a new offspring.
  Mutation: random alterations made to an individual's genetic code, leading to new variations that could potentially improve fitness.
  Iteration: the process of fitness evaluation, selection, and reproduction is repeated over multiple generations. This iterative process continues until a termination criterion is met.

The individuals in each population are called genomes and represent networks, which are composed of both nodes and connections. The neural networks in NEAT begin evolution with a first population with very simple genomes which grow over successive generations. NEAT extends these basic GA principles by incorporating:

- **Speciation** to protect innovation: the individuals in the evolving population are grouped by similarity into species, and each of them can compete only with the individuals in the same species, ensuring diverse evolutionary paths. This process, known as speciation, is guided by the compatibility distance δ:

$$\delta = c_1 \cdot \frac{E}{N} + c_2 \cdot \frac{D}{N} + c_3 \cdot \overline{W} \tag{3.3}$$

  Where δ corresponds to the compatibility distance between two genomes, E is the number of excess genes, D is the number of disjoint genes, W is the average weight difference of matching genes, N is the normalization factor (typically the number of genes in the larger genome), c_1, c_2, c_3 are coefficients weighting the importance of excess genes, disjoint genes, and weight differences, respectively.

- **Historical markings** to enable meaningful crossover between different network topologies. In traditional GAs, crossover involves exchanging genes between parents to create offspring. This process becomes challenging when dealing with vastly different structures, such as varying numbers of nodes and connections. To correctly align these genes (nodes or connections), each new gene is assigned a unique innovation number when it is created. Genes with matching innovation numbers are aligned during crossover, while those that do not match are considered disjoint or excess genes, as illustrated in Figure 3.10. The innovation number is incremented whenever a structural mutation occurs (when adding connection mutation or node mutation), allowing the tracking of gene origins across generations and ensuring proper alignment between two networks during crossover. Disjoint genes are those with different innovation numbers but that lie within the overlapping range of genes from both parents. Excess genes are those that are present in only one parent and absent in the other.

- **Incremental growth from minimal structure**, i.e. the evolution starts with simple networks and gradually increases their complexity. The algorithm begins the evolutionary process with simple neural networks, typically containing only input and output nodes with no hidden nodes, and incrementally grows their complexity through mutations. This approach prevents the evolution from starting with overly complex networks that are hard to optimize and understand. The complexity either increase by adding a node during mutation: a connection between two existing nodes is split and a new node is inserted in between, or by adding connection between two previously unconnected nodes. We will test this approach by comparing it to another case where the networks begin with a predefined random topology.

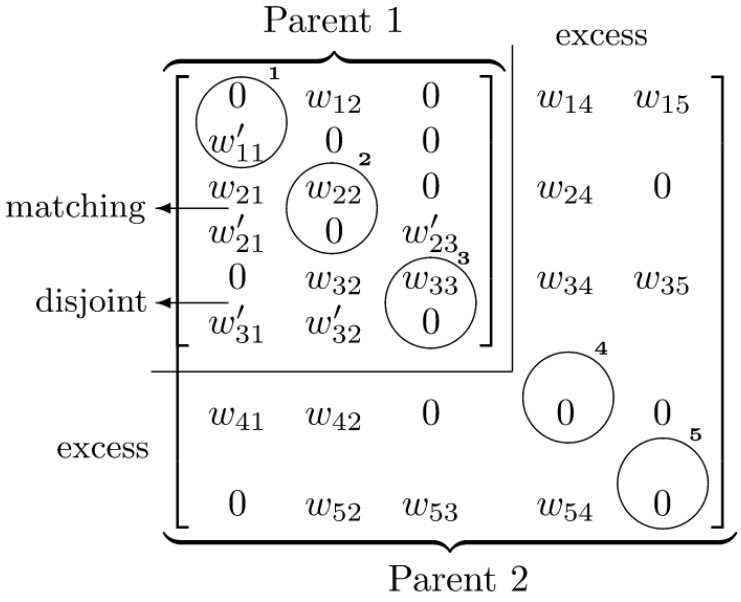

Figure 3.5: The circles represent the nodes and have an innovation number equal to their position in the diagonal of the matrix W. During the crossover, the matrices are aligned according to their innovation number. Disjoint genes are those with differing innovation numbers but that lie within the overlapping range of genes from both parents. Excess genes are those that are present in only one parent and absent in the other. (Figure taken from Chatzidimitriou and Mitkas [2010].)

3.2.3 NEAT implementation in the economic decision-making task

We use a Python implementation of NEAT called PEAS[1]. One limitation of the NEAT algorithm is its high computational and time costs. To run the simulations efficiently, we utilized a large High-Performance Computing environment known as PLAFRIM[2].

Each time a new population is generated, every genome representing a network structure is evaluated on an economic task. Specifically, the weights W and W_{in} from each genome are used to construct an ESN. The ESN then undergoes training to adjust the readout layer W_{out}. This training is performed on lotteries 2, 4 and 6 to optimize the network's performance on these tasks. Once training is complete, the structure is tested on lottery 6 to evaluate its performance. The fitness score to be maximized during the evolutionary process is calculated as the total reward cumulated during this testing phase, defined as:

[1]Peas (https://github.com/noio/peas)
[2]Plafrim (https://www.plafrim.fr/)

$$fitness = \sum rewards_{Lottery6} \tag{3.4}$$

This fitness function drives the evolution by selecting the structures that achieve the highest rewards during testing. This approach is designed to simulate survival-related tasks associated with food acquisition. The initial reservoir comprises N=200 neurons. Before executing the genetic algorithm, we first optimize the hyperparameters of a randomly initialized reservoir. The weights are initialized according to a normal distribution with the mean fixed to $\mu = 0$ and the variance of $\sigma^2 = sr^2/(rc_{connectivity}.n_{units})$. The optimized hyperparameters are shown in Table 3.1.

ESN Hyperparameter	**Value**
Input nodes	24 (2*12)
Output nodes	200
sr	0.012
lr	0.325
input_connectivity	0.2601
rc_connectivity	0.267
output_connectivity	1
η	0.003
decay	0.700

Table 3.1: ESN Hyperparameter Values: Except for the input and reservoir connectivity, all other parameters remain fixed throughout the entire evolutionary process.

With the parameters optimized, the model effectively performed the task by consistently selecting the lottery with the highest expected value, i.e., the most rewarding lottery during the testing phase. This is illustrated in Figure 3.6, which shows the probability of choosing lottery 1 as a function of the difference in expected values between lottery 1 and lottery 0. Across the 20 seeds, the sigmoid-like curves demonstrate that the model reliably chose the lottery with the greater expected value. The probability was calculated by dividing the number of times the model selected lottery 1 in a given trial by the total number of responses for that trial. This is expressed as: $P(choose_1) = \frac{n(choose_1)}{n_{total}}$.

During the evolutionary process, hyperparameters of Table 3.1 are kept fixed except for the input and reservoir connectivity that will evolve on the same time as the structure. The NEAT algorithm process begins with an initial population of 100 individuals configured with the simplest possible topology. This is achieved by setting up a fully connected network between the input and output nodes, where each input node is connected to each output node. The weights are initialized using a normal distribution with a mean of 0 and a specified initial standard deviation. The parameters required for the evolutionary process are detailed in Table 3.2. During evolution, nodes and connections undergo mutations through various operations,

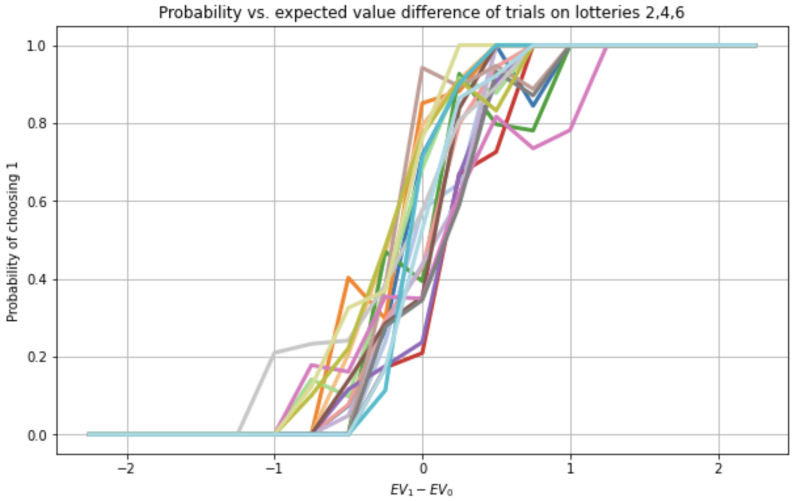

Figure 3.6: Probability of choosing lottery 1 over lottery 2 based on the difference in their expected values after model training. The model effectively selects lottery 1 when its expected value is higher compared to lottery 2, and vice versa, demonstrating its proficiency in performing the task. The model were testing on 10 different seeds corresponding to the different colors.

including adding new nodes with a certain probability, creating new connections, and modifying the properties of existing connections. These modifications include mutating weights, resetting connections, and modifying biases, with a certain probability as specified in Table 3.2. The top-performing individuals from each species are selected for crossover and mutation. Stagnant species (those showing no improvement) or species with too few members are eliminated. Elitism ensures that the best individuals are preserved across generations without any alterations, maintaining the highest-performing solutions throughout the evolutionary process.

3.3 Results

The population of the first generation is initialized in three different variants:

- **Simple Topology**: the initial population is composed of the simplest configuration: a fully connected network between input and output nodes, with no hidden nodes.
- **Random topology**: the initial population is composed of networks configured with a random topology including recurrent connections between neurons.
- **Risk-Seeking Population**: the initial population consists of networks specifically designed to include only risk-seeking individuals.

Parameter	Value
Initial weight std. deviation	0.01
Weight range	[-1, 1]
Prob. of adding node	0.1
Prob. of adding connection	0.1
Prob. of mutating weight	0.2
Std. deviation of weight mutation	0.7
Prob. of resetting weight	0.05
Prob. of mutating bias	0.2
Std. deviation of bias mutation	0.1
Prob. of re-enabling connection	0.1
Prob. of disabling connection	0.1
Distance excess (c_1)	1
Distance disjoint (c_2)	1
Distance weight (c_3)	1

Table 3.2: NEAT Algorithm Parameters

3.3.1 Initiating evolution from the simplest topology

Fitness evolution

The NEAT algorithm was executed over 100 generations comprising 100 individuals, whose structures were evaluated on the economic task. The fitness function represents the sum of rewards cumulated over 1,000 trials using lottery 6. The selection process consistently favored individuals that either improved or maintained high fitness levels as illustrated in Figure 3.7. The scores of the models are defined as the percentage of successful choices, i.e. the selection rate of the lottery with the highest expected value during the testing phase containing 1,000 trials of Lottery 6. The results are displayed in the form of boxplot in Figure 3.7. Across all generations, the interquartile range of success consistently falls between 80% and 90%, indicating that most models performed the task effectively. A t-test was conducted to compare the scores of individuals in generation 1 with the score of individuals in generation 95, yielding a p-value = 5e-08. This result indicates that the performance of individuals in generation 95 was significantly better than that of those in generation 1 and confirms that the evolutionary process successfully produced models with better performances throughout the generations.

Structural analysis

3D PCA was applied to individuals across generations, allowing to provide a visual representation of the evolutionary process as depicted in Figure 3.8. The figure reveals distinct branches that emerge over time, indicating that NEAT explores diverse species throughout generations. In Figure 3.8-left, colors indicate the generation number, highlighting the progression of the NEAT process. Over successive generations, the NEAT algorithm explores various species, with each branch hypoth-

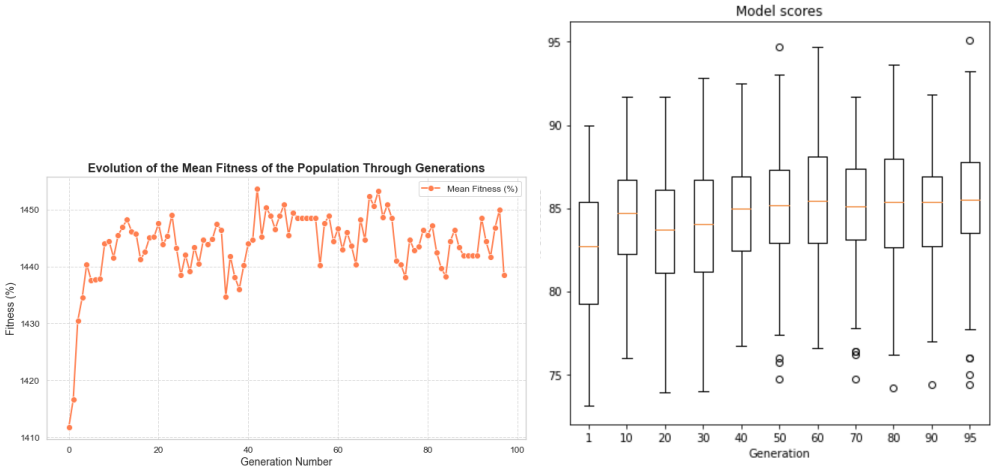

Figure 3.7: **Left:** Fitness evolution over the generations, which correspond to the cumulated rewards during the testing process containing 1000 trials of lottery 6. **Right:**Boxplots showing the scores of individual models across generations. The score represents the percentage of successful choices, defined as selecting the lottery with the highest expected value when the options differ. The interquartile range consistently falls between 80% and 90%, indicating that the models were able to perform the task effectively. A t-test was conducted to compare the scores of individuals in generation 1 with the score of individuals in generation 95, yielding a p-value = 5e-08. This result indicates that the performance of individuals in generation 95 was significantly better than that of those in generation 1 and confirms that the evolutionary process successfully produced models with better performances throughout the generations.

esized to represent a different evolutionary path. Figure 3.8-right shows the same 3D PCA analysis, but with colors corresponding to the density of each structure, calculated as follows :

$$\text{density} = \left(\frac{\text{non\_zero\_elements}}{\text{total\_elements}}\right) \times 100$$

This color coding helps trace the evolution of different species, with the center representing earlier generations characterized by simpler, less dense topologies. The branches represent distinct species, most of which are sparse (orange), while a smaller fraction (red) corresponds to denser networks.

When examining the architecture of each genome, two distinct types of networks that are visually distinguishable were identified and illustrated in Figure 3.9. The first type, referred to as the *Sparse Network*, features a pattern of sparse, random connections. The second type, known as the *Hub Network*, is characterized by a single neuron that connects to nearly all other neurons. Hub networks exhibit a higher density (density > 0.3), while sparse networks have a lower density (density

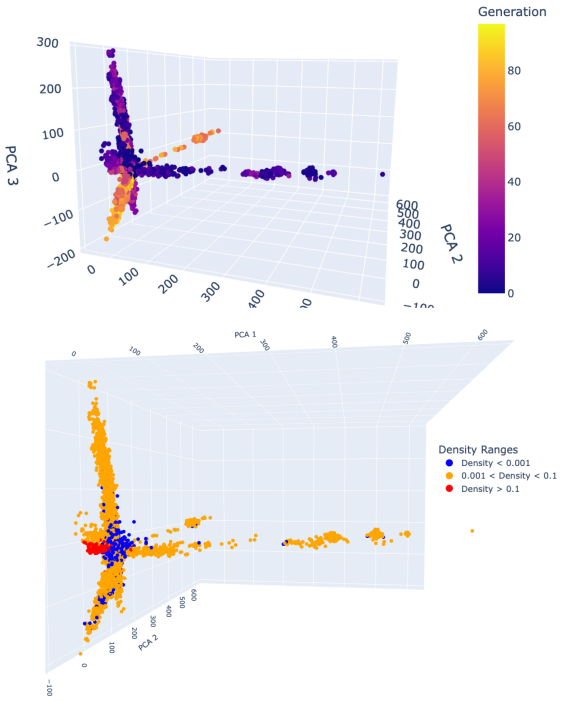

Figure 3.8: 3D PCA analysis of all individuals across generations, illustrating the evolutionary process. Distinct branches represent the exploration of different species. **Top**: Colors indicate generation numbers, allowing visualization of the NEAT algorithm's progression through generations. Each branch is hypothesized to represent a different evolutionary trajectory. **Bottom**: The same 3D PCA analysis with colors corresponding to structural density. This color scheme highlights the progression of species over time, with the center representing earlier generations with simpler, less dense structures. Branches represent different species, most of which are sparse (orange), while a small fraction (red) represents denser networks.

< 0.1). This density difference enables us to track the evolution of the two network types over time and in the PCA space as represented in Figure 3.9. The emergence of these two network types supports the degeneracy principle [Edelman and Gally, 2001], which states that different structures can achieve the same task with similar performances.

The evolution of the number of network types is illustrated in Figure 3.11-right, indicating that Hub networks form a smaller community compared to Sparse networks, which likely represent multiple species types since they appear across different branches of the 3D PCA. Hub networks emerge around generation 20 but vanish by generation 80, suggesting that this species took some time to evolve but was ultimately unable to persist. Since Hub. networks have a notably higher density, this pattern is also apparent when examining the mean density of all individuals per

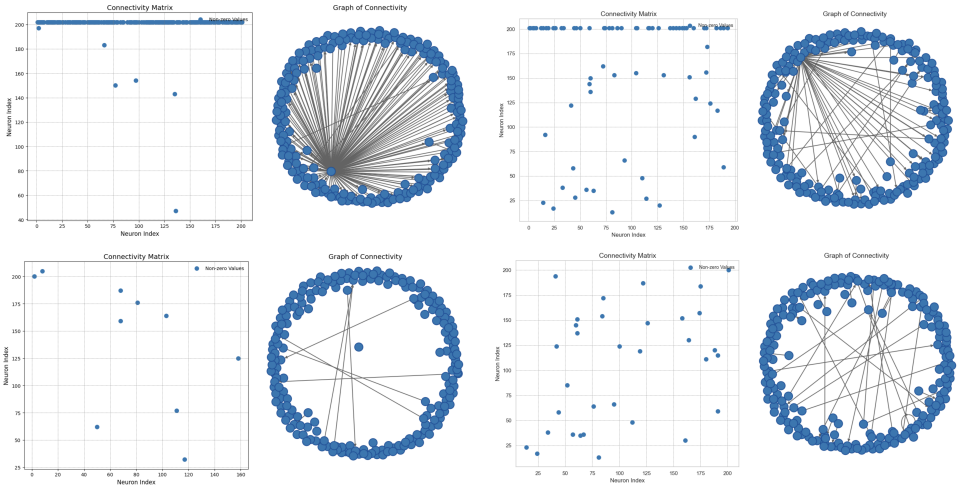

Figure 3.9: Emergence of Sparse and Hub networks from the evolutionary algorithm. **Top**: Hub networks, where the neuron indexed at 200 is highly connected to almost all other neurons, while the remaining neurons exhibit sparse connections. **Bottom**: Sparse networks, characterized by neurons that are connected to each other in a random, sparse manner. **Left**: Connectivity matrix of the weight matrix W. **Right**: Graph representation of the network generated using the Networkx Python library [Hagberg and Conway, 2020], visualized as directed graphs with self-loops.

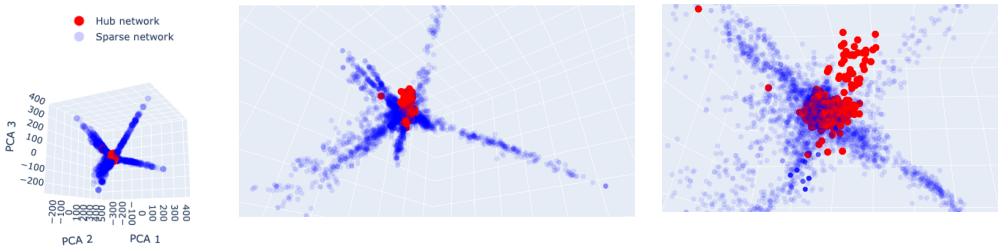

Figure 3.10: 3D PCA applied on every individual of all generation. The colormap corresponds to the network type. From left to right: zoom of the same caption.

generation, as shown in Figure 3.11-left. If the Hub community fails to persist, it suggests that they were not earning sufficient rewards to be selected for the next generation. Figure 3.12 illustrates this by showing the mean fitness of each network type per generation, which reflects the total reward received during the testing phase. Hub networks consistently receive slightly fewer rewards on average and from generation 80, their cumulative reward is too low to secure selection for the next generation. A t-test showed that the comparison of cumulated rewards between the two simulations are not significantly different, however, this seems enough to allow to the survival of one species and not the other.

Despite sharing the same hyperparameters, the only distinguishing factor between these networks is their structure. Consequently, two distinct structures emerge from the evolutionary process, highlighting degeneracy; however, they exhibit differences in performance when averaged at the population level. This suggests that network structure may have an influence performance in our economic decision-making task.

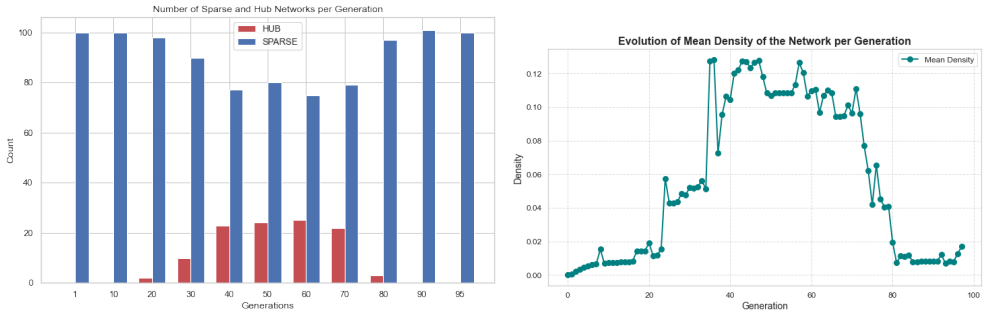

Figure 3.11: **Left**: number of Hub and Sparse networks per generation. Hub networks emerge starting from generation 20 but do not persist beyond generation 80. **Right**: Mean density per generation. The mean density sharply increases after generation 20 with the emergence of Hub networks, followed by a sudden decrease due to their disappearance.

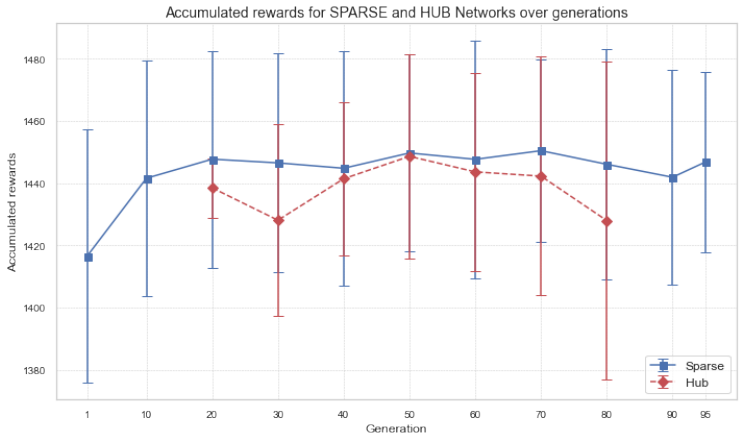

Figure 3.12: Mean cumulated rewards (fitness of the evolutionary algorithm) per generations according to the network type. The Hub networks do not accumulate a sufficient amount of rewards to survive.

Risk-taking behavior analysis

We hypothesize that the observed difference in cumulated rewards between the two networks may come from variations in risk-taking behavior. To investigate this hypothesis, we focused on trials where the expected values of the two lotteries are equal, with one lottery offering higher risk but potentially greater rewards, and the other providing a safer option with lower rewards. We analyzed the percentage of safe choices made by each model in these equal-expected-value trials. Our interpretation framework was as follows:

- Risk-neutral individuals: expected to choose the safe option approximately 50% of the time.
- Risk-seeking individuals: expected to choose the safe option less than 50% of the time.
- Risk-averse individuals: expected to choose the safe option more than 50% of the time.

Figure 3.13 illustrates the percentage of safe choices made during trials with equal expected value, with Sparse networks shown on the left and Hub networks on the right. Each point represents the behavior of one individual per generation. The red dotted line marks a 50% preference for the safest choice, indicating risk-neutral behavior. Individuals below this line are characterized as risk-seekers, while those above it are identified as risk-averse. In both networks, a higher proportion of individuals demonstrated risk-averse behavior, consistently choosing the safe option more frequently. This finding aligns with previous studies on risk-taking behavior, which suggest that humans [Kahneman and Tversky, 1979] and other species such as rodents [Constantinople et al., 2019], birds [Caraco, 1981] and insects [Harder and Real, 1987], , plants and macaques [Nioche et al., 2021] tend to favor risk-averse decisions in gain scenarios.

Nevertheless, a fraction of individuals in each generation showed risk-seeking tendencies, choosing the safe option less than 50% of the time, indicating variability in risk preferences across generations. In the Sparse network community, a significant proportion of risk-seeking individuals is maintained. The proportion is significantly smaller in the Hub network community. A t-test was conducted to compare the percentage of risk-seekers between the two network structures. The resulting p-value of 2.26e-5 indicates a significantly higher proportion of risk-seekers in the Sparse network compared to the Hub network. This substantial difference between the Sparse and Hub species may provide insight into their differential survival rates. We hypothesize that the Hub species' lack of sufficient risk-seeking individuals may affect its ability to drive evolution and maintain species survival. This aligns with the perspective of Trimpop [1994] who argue that risk-taking behavior is essential to evolution, i.e. some level of risk-taking is necessary for

species to adapt and survive in changing environments.

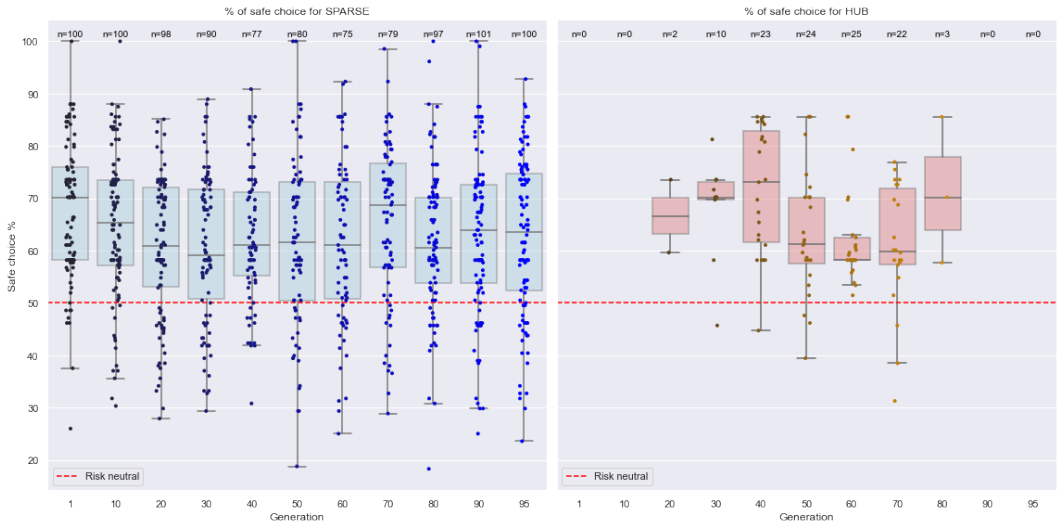

Figure 3.13: Percentage of safest choices made during trials with equal expected values, with Sparse networks displayed on the left (in blue) and Hub networks on the right (in red). Each point represents an individual per generation, and the boxes illustrate the boxplot for each generation. The red dotted line marks a 50% preference for the safest choice, indicating risk-neutral behavior. Individuals below this line are categorized as risk-seekers, while those above are considered risk-averse. Over generations, a higher proportion of individuals exhibit risk-averse behavior. A t-test yielded a p-value of 2.26e-5, indicating a significantly higher proportion of risk-seekers in the Sparse network compared to the Hub network.

3.3.2 Initiating evolution from a complex random topology

In the previous subsection, we started the evolutionary process with the simplest possible network topology (no recurrent connections in the reservoir, and fully connected input and output connections) and gradually increased complexity over successive generations. We now consider whether the same evolutionary effects would emerge if we began with an already complex Sparse network. How the network would evolve, what it would evolve towards, the potential development of alternative structural solutions, and the impact on risk-taking behavior. To explore these questions, we replicated the same evolutionary process, but instead of starting with the simplest topology, we begin with a randomly connected network. The NEAT algorithm starts with an initial population of 100 individuals, each configured with a random topology of 200 units, an input connectivity of 0.7, and a reservoir connectivity of 0.001. This initial configuration resembles the Sparse network defined in the previous subsection. Aside from this, the process follows the same steps as

described earlier, with the same ESN and NEAT parameters described in Table 3.1 and Table 3.2.

Fitness evolution

The evolutionary process successfully produced models with good performance throughout the generations, favoring individuals that either improved high fitness levels, high score performances as illustrated in Figure 3.14. A t-test was conducted to compare the scores of individuals in generation 5 with the score of individuals in generation 90, yielding a p-value = 8.1e-10. This result indicates that the performance of individuals in generation 90 was significantly better than that of those in generation 5 and confirms that the evolutionary process successfully produced models with better performances throughout the generations.

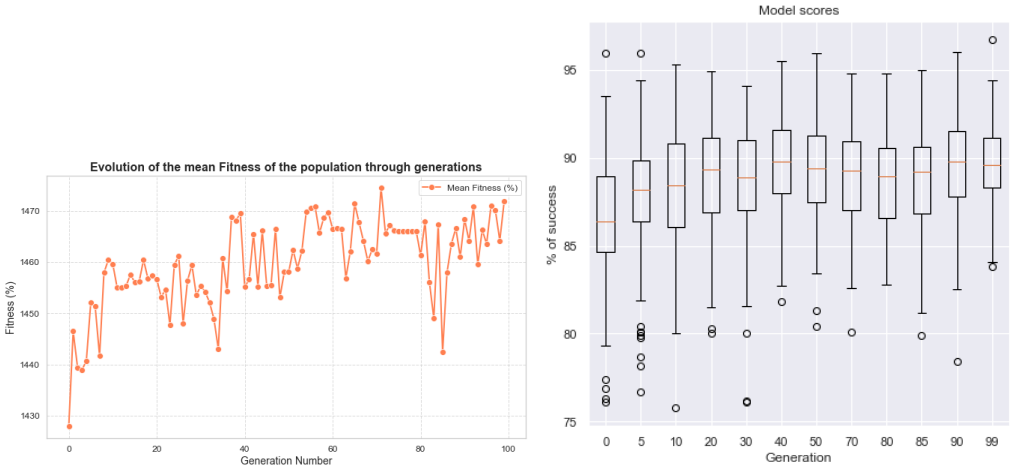

Figure 3.14: **Left:**Fitness evolution over the generations, when starting with a random topology. **Right:**Boxplots showing the scores of individual models across generations. The interquartile range consistently falls between 80% and 97%, indicating that the models were able to perform the task effectively.

Structural analysis

The obtained 3D PCA is composed of different branches illustrating the evolutionary process, similar to those presented in the previous section and therefore omitted here to avoid redundancy. Figure 3.15 illustrates that the Hub species emerged from the population as early as generation 5 but disappeared after generation 87. This pattern is also reflected in the mean density of each generation, which shows a sharp increase corresponding to the emergence of Hub networks, followed by a sudden drop marking the species' disappearance. When starting from a random topology, the Hub species appear earlier and survivre longer.

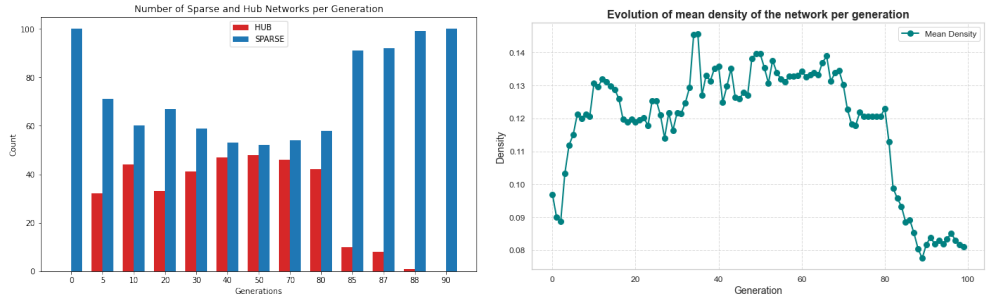

Figure 3.15: Number of Hub and Sparse networks per generation. Hub networks emerge starting from generation 20 but do not persist beyond generation 80. Since Hub networks have a notably higher density, their emergence can also be observed by looking at the mean density of each generation.

Risk-taking behavior analysis

Figure 3.16 reveals that the Sparse species exhibits a balanced mix of risk-seeking and risk-averse individuals, with the median values in the boxplots centered around 50% for choosing the safe option. This suggests an equal distribution of risk preferences within the Sparse species. In contrast, the Hub species predominantly consists of risk-averse individuals, with only a few exhibiting risk-seeking behavior. A t-test was conducted to compare the percentage of risk-seekers between the two network structures. The resulting p-value of 4.9e-4 indicates a significantly higher proportion of risk-seekers in the Sparse network compared to the Hub network. Thus, consistent with earlier findings, there is a notable asymmetry in risk-taking behavior between the Sparse and Hub networks. Nevertheless, the Hub species survive during a bigger number of generations compared to the previous results (70 generations against more than 90 generations).

Figure 3.12 presents the mean cumulative rewards of each network type across generations. Hub networks consistently earned slightly more rewards during their existence than Sparse networks, indicating that being risk-averse during this period was advantageous compared to being excessively risk-seeking. It appears that having too many risk-seekers led to greater risks without being lucky, resulting in lower overall gains. In contrast, a predominantly risk-averse strategy proved beneficial—up to a point. Around generation 70, the trend reversed: Hub networks began to receive fewer rewards on average, while risk-seeking behavior became increasingly advantageous. This shift continued until generation 88, when the Hub community's cumulative rewards dropped too low for the species to survive. Thus, consistent with earlier findings, Hub networks may have failed to survive longer due to their lack of sufficient risk-seeking individuals, which limited their ability to maintain high cumulative rewards and, ultimately, the species' survival.

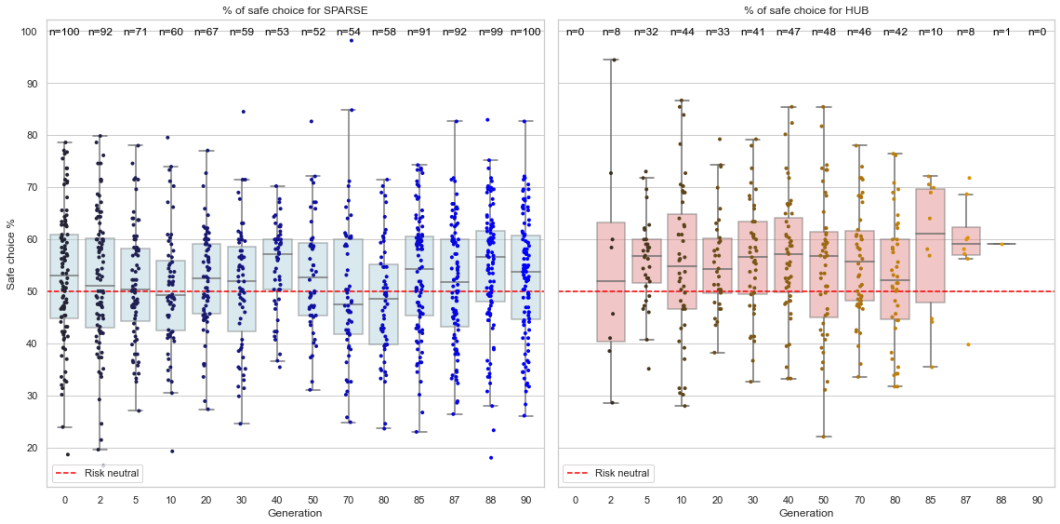

Figure 3.16: Percentage of safest choices made during trials with equal expected values, with Sparse networks displayed on the left (in blue) and Hub networks on the right (in red). Each point represents an individual per generation, and the boxes illustrate the boxplot for each generation. Over generations, a higher proportion of individuals exhibit risk-averse behavior. A t-test yielded a p-value of 4.9e-4, indicating a significantly higher proportion of risk-seekers in the Sparse network compared to the Hub network.

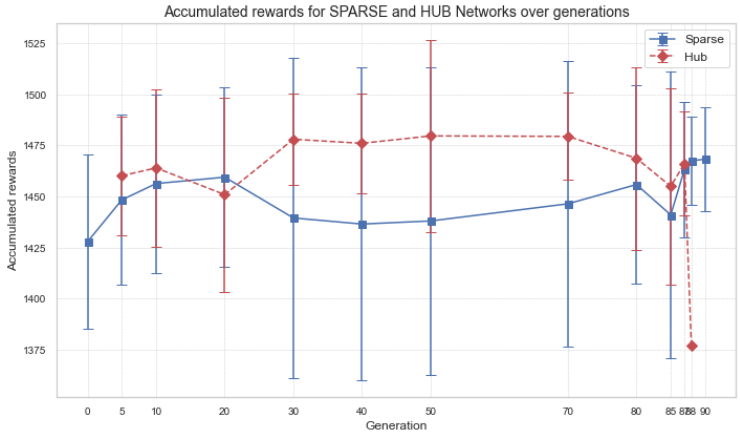

Figure 3.17: Mean cumulated rewards (fitness of the evolutionary algorithm) per generations according to the network type. The Hub networks do not accumulate a sufficient amount of rewards to survive.

3.3.3 Initiating evolution from a population biased toward risk-seeking behavior

In the previous two subsections, we initiated the evolutionary process from both the simplest possible topology and an already complex network. In both cases, the

results converged: two distinct networks emerged, characterized by an asymmetry in risk-taking behavior. The Hub networks were predominantly risk-averse, while the Sparse networks contained a higher proportion of risk-seekers, resulting in a more balanced population overall and leading to improved survival chance. These results strongly suggest that the presence of risk-seeker individuals are essential to the survival of the species.

In this final subsection, we shift our focus to an evolutionary process starting with a population composed exclusively of risk-seeker individuals. To achieve this, we biased the reservoir models during hyperparameter optimization. During the training phase, when models faced trials with equal expected values, the probability of receiving a reward was artificially set to 1 if the model selected the riskier option (i.e., the option with a smaller probability of success). This approach was designed to encourage the model to favor the risky choice. As a result, the readout layer (W_{out}) was trained to predominantly select the riskiest option when faced with equal expected values, thereby instilling a risk-seeker behavior. This readout layer was then frozen and kept constant across all structures throughout the evolutionary process. Unlike the previous cases, where individuals were retrained in each generation, here all individuals retained the same W_{out} throughout the evolution.

Fitness evolution

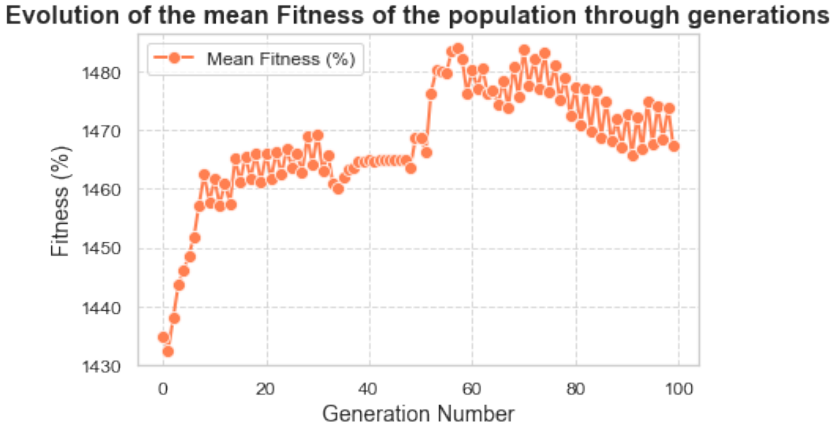

Figure 3.18: Fitness evolution over the generations on the risk-biased models.

The optimized hyperparameters are show in Table 3.3. With these parameters, the model effectively performed the task by consistently selecting the lottery with the highest expected value, i.e., the most rewarding lottery during the testing phase and select more often the risky option when equal expected values. The NEAT algorithm process begins with an initial population of 100 individuals configured with a sparse connectivity. During evolution, unlike in the previous sections, only the weights evolve, while the nodes and connections remain unchanged. The probabilities of

adding nodes and connections were set to zero, as detailed in Table 3.4. This decision was made for simplicity, as adding nodes to the reservoir would be impractical given that the readout layer is kept fixed. Consequently, since the network structure does not evolve, we do not expect to see the emergence of distinct Hub and Sparse networks.

ESN Hyperparameter	**Value**
Input nodes	24 (2*12)
Output nodes	200
sr	0.003
lr	0.438
input_connectivity	0.34
rc_connectivity	0.004
output_connectivity	1
η	0.003
decay	0.545

Table 3.3: Parameter Values

Parameter	**Value**
Initial weight std. deviation	0.01
Weight range	[-1, 1]
Prob. of adding node	0
Prob. of adding connection	0
Prob. of mutating weight	0.1
Std. deviation of weight mutation	0.1
Prob. of resetting weight	0.02
Prob. of mutating bias	0
Std. deviation of bias mutation	0
Prob. of re-enabling connection	0
Prob. of disabling connection	0
Distance excess (c_1)	1
Distance disjoint (c_2)	1
Distance weight (c_3)	1

Table 3.4: NEAT Algorithm Parameters

Structural analysis

The selection process consistently favored individuals that improved fitness levels, as illustrated in Figure 3.19. The application of 3D PCA to individuals across generations, shown in Figure 3.19, reveals distinct branches emerging over time, indicating that NEAT explores a diverse range of species throughout generations.

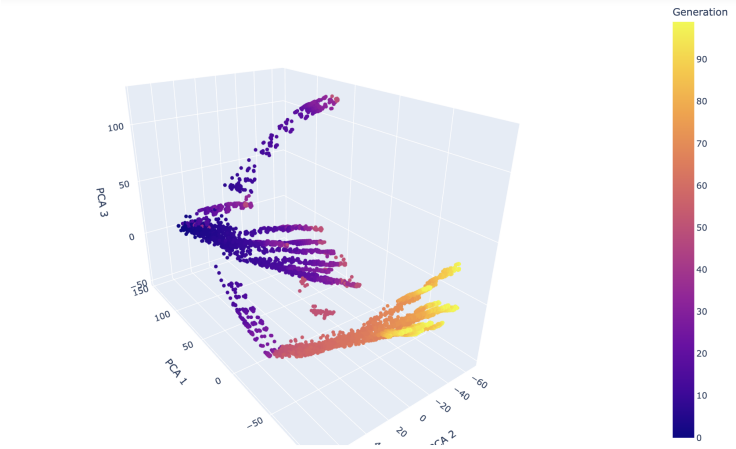

Figure 3.19: 3D PCA analysis of all individuals across generations, illustrating the evolutionary process.

Risk-taking behavior analysis

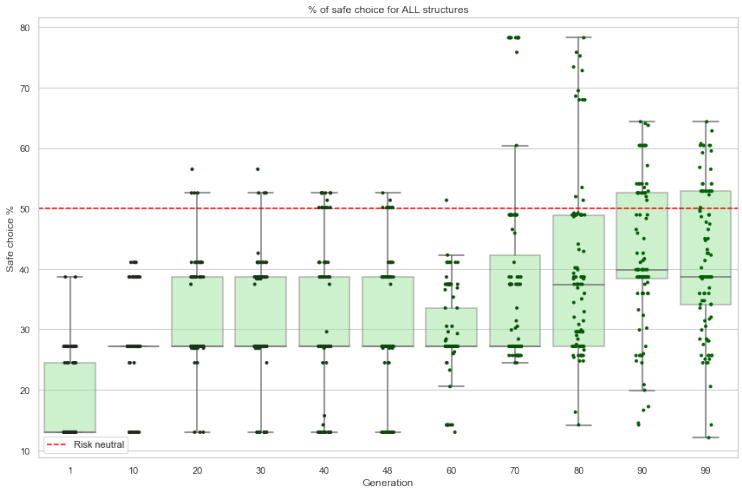

Figure 3.20: Percentage of safest choices made during trials with equal expected values. Each point represents an individual per generation, and the boxes illustrate the boxplot for each generation. Initially, the populations exhibit only risk-averse individuals, consistent with the fact that the starting population was biased. Risk-averse individuals began to appear from generation 20, with their numbers steadily increasing.

Figure 3.20 displays the mean percentage of safe choices per population over generations. Initially, the populations exhibit highly risk-averse behavior, consistent with the fact that the starting population was biased. Interestingly, despite all individuals being composed of a risk-seeker readout, risk-averse individuals began to appear from generation 20, with their numbers steadily increasing. This emergence

of risk-averse individuals is unexpected given that they possess risk-seeking bias of the readout layer. This suggests that the evolutionary algorithm adapts the network topology to counterbalance the risk-seeking tendencies of the readout, leading to a more risk-averse population, strongly suggesting that risk-averse individuals also play a significant role in driving evolution.
While previous results indicated that risk-seeking individuals are crucial for the species' survival, these findings strongly demonstrate that risk-averse individuals are also essential for advancing evolution.

3.4 Discussion

By evolving the structure of a reservoir network on an economic decision-making task, we showed that starting from the simplest structure, the evolutionary process evolve towards two types of structures, with one emerging straight from the beginning- the Sparse network, and a second one emerging some generations later- the Hub network. Compared to the Sparse population, the Hub population does not survive after several generations. To explore this further, we repeated the experiment, starting with a population of Sparse individuals rather than the simplest topology. Despite this initial bias towards the Sparse network type, the Hub species emerge within the population, and vanishes after some generations. These initial results indicate that, in both evolutionary configurations, at least two distinct species emerged, suggesting that various network structures can solve the same task. This observation supports the concept of multiple realizability and degeneracy: different structures can lead to similar behaviors [Edelman and Gally, 2001].

Both evolutionary configurations revealed similar tendencies in risk-taking behavior: although both species performed comparably on the task, they differed notably when confronted with lotteries of equal expected value. The Hub population was predominantly composed of risk-averse individuals, while the Sparse population also included a significant proportion of risk-seekers. We hypothesize that this difference may explain why the Hub species struggles to survive: while risk-aversion can be advantageous for a time, over successive generations, the lack of sufficient risk-taking may ultimately affect the species' ability to maximize cumulative rewards and adapt effectively to the environment.

At the individual level, both species exhibited a mix of risk-seeking and risk-averse behaviors, highlighting the degeneracy aspect of the problem. However, the overall tendencies differed at the population level, underscoring that the structure of decision-making networks can influence risk-taking behavior.

These findings align with research thesis of Perez Rodero, which demonstrated differential brain connectivity patterns for conditions related to risk perception and

risk behavior. Further support comes from Deza Araujo et al. [2018] who investigated the neural correlates of risk-seeking behavior, particularly in loss scenarios. Their findings revealed that higher risk-seeking tendencies for losses were associated with increased connectivity between medial temporal regions, frontal regions, and the Default Mode Network (DMN). The observation suggests that the structure and connectivity of the networks appear to play a crucial role in shaping risk preferences.

Furthermore, our results suggest that certain behaviors are crucial for evolutionary success. The first two experiments highlight the necessity of a substantial proportion of risk-seekers for species survival. The final experiment further illustrates that while the presence of a fraction of risk-averse individuals may be essential, a population composed solely of risk-seekers might also fail to survive, and risk-averse behavior is also crucial for maintaining the species. This suggests that evolutionary processes may favor a balance between risk-seeking and risk-averse behaviors.

The relationship between risk-taking behavior and evolutionary advantage has been explored in various studies. Rather than evolving the structure or incorporating learning mechanisms, Nioche et al. [2021] conducted a study where they evolved directly the parameters controlling risk-aversion in agents. Their evolutionary simulations demonstrated a tendency towards risk-seeking behaviors. Interestingly, when comparing these results with macaque behaviors, they found consistency only in loss scenarios, suggesting that risk-seeking behavior may be evolutionarily advantageous in situations involving potential losses, reflecting decision-making strategies in life-threatening situations. Testori et al. [2022]'s results suggest that the combination of selfishness and risk-seeking traits can be evolutionarily advantageous for individuals, more than the combination of generosity and risk averse traits. They also observed that communities with a higher proportion of selfish risk-seeking agents could grow to larger population sizes.

The concept of risk sensitivity as an adaptive trait was further explored by Hintze et al. [2015]. Their research provides an evolutionary explanation for the risk-averse behavior observed in humans. They found that risk sensitivity evolved as a preferred strategy particularly in small groups, suggesting that small group dynamics could be driving the evolution of risk sensitivity. Homma and Takezawa [2024] focused on the role of reinforcement learning in the evolution of risk preferences. They found that learning rates are shaped by evolutionary processes, leading to domain-specific risk preferences. Notably, their evolved agents exhibited behavioral patterns consistent with prospect theory when faced with risky options of identical expected values. Finally, Kolodny and Stern [2017] emphasized that the evolution of risk preferences is a complex phenomenon influenced by multiple factors that have to be taken in consideration including the population size, the number of risky decisions in an individual's life history in shaping risk-taking

behavior and reproduction dynamics.

The nature of risk-taking behavior has been a subject of several research with evidence supporting both genetic and environmental influences [Bell, 2009]. Karlsson Linnér et al. [2019] conducted a genome-wide association study that identified hundreds of genetic loci associated with general risk tolerance and specific risky behaviors, providing evidence for a genetic component in risk-taking behavior. Aydogan et al. [2021] developed a risk score for risky behavior using genome-wide association study data from an independent sample of 297,025 individuals and provide evidence that neuroanatomical structure forms the basis for mechanisms underlying individual differences in risky behaviors. Nevertheless, it has also been shown that environmental influences play also a crucial role in shaping risk-taking behavior. Josef et al. [2016] emphasized that these preferences can evolve due to environmental factors like experience and context. [Stern, 2010] introduced the concept of "cultural inheritance" as a significant factor in the evolution of risk-taking behavior. Zhang et al. [2014] proposed that risk aversion emerges as an adaptive trait in environments where reproductive risk is systematic, i.e., correlated across individuals in a generation, suggesting that risk-taking behavior is not only determined by genes but also by the adaptive responses to specific environmental conditions. These findings highlight the complex relationship between genetics and environmental factors with risk-taking behavior.

The next chapter delves deeper into the following question: to what extent is risk-taking behavior genetically determined, and how is it shaped by environmental factors, including experience and context? To address this question, we will study the behavior of a community of monkeys. Our investigation begins with the hypothesis that environmental factors, particularly social dynamics, may play a significant role in modulating risk-taking behavior. Specifically, we aim to explore whether social hierarchy within the monkey community influences individual risk-taking propensities.

4 Chapter 4: Influences Beyond Structure: Analyzing Risk-Taking Behavior and Social Factors in Monkeys

In the previous chapter, we explored how network structure might influence risk-taking behavior from an evolutionary perspective. Nevertheless, environmental factors are also thought to be important in shaping this behavior, and these factors can evolve over time [Josef et al., 2016]. In this chapter, we shift our focus to potential external influences to address the question: what additional factors might affect risk-taking behavior in economic decisions?

To address this question, we set aside the structural analysis of minimal ESN models and instead analyze real data from monkeys. We examined a large dataset involving 24 Tonkean macaques performing the economic task previously described[1]. Our analysis is based on the Prospect Theory (PT), a framework introduced by Kahneman and Tversky [1979] that has become a cornerstone for understanding behavior under risk. PT highlights several cognitive biases in humans, such as loss aversion, risk aversion in gain scenarios, and distortion of the probability perception, which we will explain and apply to interpret the macaque data.

Our findings confirm that on average, the behavior of these monkeys can be also understood within the framework of the PT, exhibiting similar cognitive biases as humans. Nevertheless, while the overall behavior aligns with the theory, we observed a significant inter-individual variability: some monkeys were more or less risk-averse, more or less loss-aversive and varied in the way they perceive probabilities. We hypothesize that social factors, particularly social hierarchy, may influence risk-taking behavior among the monkeys, but also that risk taking behavior are not static and can evolve over time.

We will begin by explain the task setup for the monkeys, followed by an introduction to the PT framework. Next, we will demonstrate that cognitive biases identified in humans are also observable in the monkey data. Finally, we will ex-

[1]This work was conducted in collaboration with Dr. Sebastien Ballesta along with his Ph.D. student Ayrton Guerillon who greatly contributed to the data analysis.

plore how inter-individual variability in risk-taking behavior may be influenced by the social hierarchy among the monkeys.

4.1 Assessing the risk-taking behavior of monkeys

4.1.1 The original monkey task

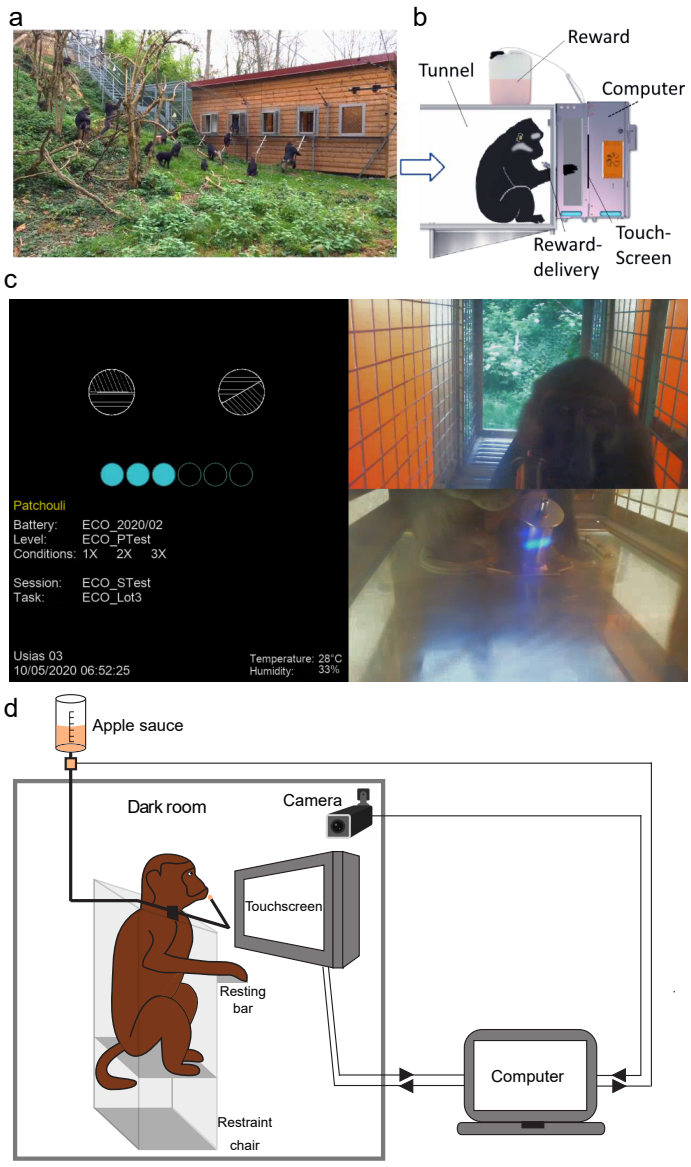

Figure 4.1: **Monkey experiment** a. Tonkean macaque's wooded park in semi-free-ranging conditions. b. cognitive tasks presented via a touchscreen interface MALT (Machine for Automated Learning and Testing) [Fizet et al., 2017]. c. Visualization of the touchscreen. d. Overall setup. (The figure is modified from Nioche et al. [2021].)

At the Centre de Primatologie (CdP) of the University of Strasbourg [Nioche et al., 2021], a social group of 24 Tonkean macaques live in a semi-free ranging conditions in a wooded park of 3788 m^2 where they could have access to a shelter at any time, as depicted in Figure 4.1-a. Inside this shelter, monkeys can play to a cognitive task via a touchscreen interface shown in Figure 4.1-b-c. The task is the same as in Nioche et al. [2021, 2019] , well detailed in section 3.1.1 of Chapter 3.

4.1.2 Data collection

	date	heure_debut	subject_id	task_id	V_left	P_left	V_right	P_right	reward	response	RT	ST	gain	loss
0	2020-02-04	07:51:07	Alaryc	1	3	1.00	-3	1.00	3.0	0.0	2101	446	3.0	0.0
1	2020-02-04	07:51:17	Alaryc	3	-3	0.75	-1	0.75	-3.0	0.0	1985	723	0.0	-3.0
2	2020-02-04	07:51:27	Alaryc	4	3	1.00	3	0.75	3.0	0.0	2190	884	3.0	0.0
3	2020-02-04	07:51:37	Alaryc	1	-3	0.25	1	0.25	0.0	0.0	2380	417	0.0	0.0
4	2020-02-04	07:51:47	Alaryc	14	-3	1.00	3	0.75	3.0	1.0	1390	309	3.0	0.0
5	2020-02-04	07:51:55	Alaryc	3	-3	0.25	-1	0.25	0.0	0.0	1449	397	0.0	0.0
6	2020-02-04	07:52:04	Alaryc	3	-3	0.75	-2	0.75	-3.0	0.0	1190	361	0.0	-3.0
7	2020-02-04	07:52:13	Alaryc	15	0	1.00	-1	0.25	0.0	0.0	909	462	0.0	0.0
8	2020-02-04	07:52:21	Alaryc	17	-1	0.50	-3	0.75	0.0	1.0	1230	246	0.0	0.0
9	2020-02-04	07:52:30	Alaryc	14	-3	1.00	3	0.50	-3.0	0.0	991	374	0.0	-3.0

Figure 4.2: **Dataset** contains behavioral information of 1 843 143 trials done by 24 monkeys. Each row corresponds to one trial during a period of three years.

Data were collected thanks to Machines for Automated Learning and Testing (MALTs) [Fizet et al., 2017]. All trials are stored in the .csv dataset represented in Figure 4.2. The columns definitions are shown in Table 4.1 and the total number of trials per monkey and per trial type is depicted in Figure 4.3.

Name	Type	Description
subject_id	string	Identification of the subject
date	datetime	Date when the trial was conducted
task_id	integer	Identification of the task (ranging from 1 to 7)
P_left	float	Reward probability for the left stimulus
V_left	float	Reward amount for the left stimulus
P_right	float	Reward probability for the right stimulus
V_right	float	Reward amount for the right stimulus
response	integer	Response (0: left, 1: right)
reward	integer	Reward delivered (1) or not (0)
RT	integer	Response time (ms)
gain	integer	Quantity of gain received at the end of the trial
loss	integer	Quantity of loss received at the end of the trial

Table 4.1: Column definitions of the behavioral dataset.

Lottery	L1	L2	L3	L4	L5	L6	L7
ALA	31228	24697	24530	24775	24380	24219	24398
OLL	15016	11726	11733	11720	11452	11314	11443
BAR	40121	31003	31167	31023	30867	30307	30823
NEM	24275	18807	18784	18790	18680	18577	18627
ALV	16900	13206	13038	13217	12796	12616	12820
CES	24363	19057	19075	19094	19009	18940	18998
ABR	25715	19779	19841	19829	19824	19539	19871
YOH	16344	12511	12427	12503	12505	12481	12491
ANU	5442	4270	4247	4274	4228	4271	4228
OLA	1746	54	55	53	53	49	47
PAC	11934	9015	8992	9088	9047	8996	9015
NER	12918	8810	8791	8843	8745	8748	8837
YIN	3437	1419	1413	1403	1406	1393	1408
DOR	14975	11327	11261	11386	11273	11366	11245
LAS	9375	6947	6897	6968	6931	6983	6935
ERI	23005	17408	17298	17416	17352	17445	17385
FIC	11790	8958	8931	8977	8884	8983	8919
OLG	3192	999	968	925	961	999	966
JEA	461	148	160	132	161	149	169
BER	5371	4032	4010	4030	4034	4018	4024
PAT	1724	1258	1253	1267	1259	1265	1249
HER	3556	2671	2663	2676	2669	2683	2675
HOR	4502	3417	3408	3417	3402	3428	3401
GAN	362	219	230	222	224	218	222

Figure 4.3: **Number of trials per monkey and per lottery type**. The control trials L1 consist of one lottery with potential gains and another with potential losses. In contrast, the trials in (L2, L4) and (L3, L5) are only composed of lotteries with potential gains or losses, respectively. The lotteries in L6 and L7 allow us to assess the risk-taking behavior of monkeys in gain and loss domains, repecitvely. For a detailed explanation of the lottery types, refer to Section 3.1.1 and Nioche et al. [2019] for further information.

4.1.3 Introduction to the prospect theory (PT)

When facing uncertainty, rational behavior is defined as making choices that maximize the expected value (EV) [Von Neumann and Morgenstern, 1947], which can be expressed as follows:

$$EV = \sum_i P_i.x_i$$

where:

- EV represents the expected value.
- x_i is the outcome,
- P_i is the probability of outcome x_i,

However, individual behavior frequently deviates from this rational ideal. One example is when asking people choosing between two situations: in the first one, they receive 100$ after flipping a coin and landing on tails, in the second one, they take home 50$ immediately. Both options have identical expected values : $0.5 * 100 = 50$ in the first case, and $1 * 50 = 50$ in the second. Thus, one would expect rational individuals to be indifferent between the two choices. However in this example, people would prefer to choose the second option, which is safest. People are not entirely logical decision-makers; they exhibit cognitive biases like risk aversion and probability distortion, leading to make choices that differ from the ones coming from the utility maximization. Prospect theory (PT) introduced by Tversky and Kahneman [1992], Kahneman and Tversky [1979, 2013] offers a framework to define this attitude toward risk. It relies on two key concepts:

- The subjective utility.
- The probability distortion.

These functions enable to describe each individual's attitude towards risk. The subjective utility function $u(x)$ reflects how each individuals perceive and value different outcomes, by assigning a numerical value to different outcomes x. In the PT framework and as depicted in Figure 4.4, the utility function is described to be concave in the gain domain, indicating that individuals exhibit risk aversion when facing potential gains. Conversely, in the domain of losses, the utility function is convexe, suggesting that humans are risk-seeker when facing potential losses. Furthermore, the utility function exhibits an asymmetry between the gain and loss domains, with a steeper curve near the reference point in the loss domain compared to the gain domain, showing that individuals are more sensitive to losses than to gains. The latter bias is named loss aversion.
The probability distortion $w(p)$ describes how people perceive and distort probability p in decision under risk. As depicted in Figure 4.4, the function has an inverted S-shape, meaning that individuals tend to overestimate small probabilities,

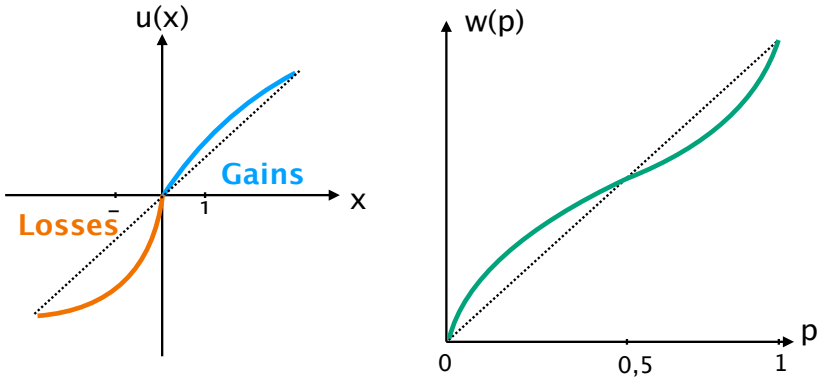

Figure 4.4: **Graphical representation of the Prospect Theory of Kahneman and Tversky [1979]. Left**: Subjective utility function. The concavity of the curve for gains indicates risk-aversion, while the convexity of the curve for losses indicates risk seeking. Loss aversion is indicated by a steeper curve for losses than for gain. **Right**: Probability distortion function: small probabilities are over-estimated while high-probabilities are under-estimated.

and under estimates big probabilities, influencing the decision-making process and choice [Tversky and Kahneman, 1992, Bruhin et al., 2010]. However, this inverted S-shape is not a consistent feature. Other studies showed different shapes of distortion [Ferrari-Toniolo et al., 2019, Farashahi et al., 2018].

The subjective utility function

In Tversky and Kahneman [1992], the subjective utility function is defined as:

$$U(x) = \begin{cases} x^{\rho^+} & \text{if } x \geq 0 \\ -\lambda(-x)^{\rho^-} & \text{if } x < 0 \end{cases}$$

where:

- $U(x)$ represents the subjective utility of a quantity x,
- λ is the loss aversion parameter. It controls the steepness of the curve in the loss domain, measuring the psychological impact of losses compared to gains. When $\lambda = 1$, gains and losses are weighted the same. When $\lambda > 1$, which is shown to be the most common behavior, people overweight losses compared to gain, making them loss-aversive. Conversely when $\lambda < 1$, people underweight losses compared to gains.
- ρ^+ and $\rho-$ are the risk aversion parameters. They control the concavity and the convexity of the curve. If $\rho^{+/-} = 1$, $U(x) = x$, indicating risk-neutral preferences. If $\rho^{+/-} < 1$, the curve is concave, indicating risk-aversive

behavior. If $\rho^{+/-} > 1$, the curve is convexe, indicating risk-seeking behavior. We will use the expression for the rest of our analysis.

Since the development of the PT, the existing literature has been indicating that the shape of probability distortion varies across decision-making contexts, experimental settings, and individuals. Numerous formulations of probability distortion have been proposed, but no single unifying model has emerged[Zhang et al., 2020]. Similarly to Tymula et al. [2023], we gathered each of these expressions:

- **TK** (from Tversky and Kahneman [1992])

$$w(p) = \frac{p}{(p^\alpha + (1-p)^\alpha)^{\frac{1}{\alpha}}} \quad \text{for } 0 < \alpha \leq 1$$

  α represents the curvature/distortion of the function. if $\alpha = 1$, the function is linear i.e. $w(p) = p$. If $\alpha < 1$, the function is inverse S-shaped, indicating that the function overweight small probabilities and underweight big probabilities.

- **P1** (from Prelec [1998]):

$$w(p) = \exp(-(-\ln p)^\alpha), \quad \alpha > 0$$

  α represents the curvature/distortion of the function. If $\alpha = 1$, the function is linear i.e. $w(p) = p$. If $\alpha < 1$ the function overweights small probabilities and underweights big probabilities. if $\alpha > 1$, it's the contrary effect.

- **P2** (from Prelec [1998]) :

$$w(p) = \exp(-\delta(-\ln p)^\alpha), \quad \alpha > 0, \quad \delta > 0$$

  α represents the curvature/distortion of the function and has the same interpretation as in P1 expression.

- **GE** (from Goldstein and Einhorn [1987]) :

$$w(p) = \frac{\delta p^\alpha}{\delta p^\alpha + (1-p)^\alpha}, \quad \delta > 0, \quad , \alpha > 0$$

  δ represents the sensibility/scaling of the probability p to the probability distortion function. If $\delta > 1$, the function becomes more sensitive to changes in the probability, if $\delta < 1$, the function becomes less sensitive. α corresponds to the curvature or the function. If $\alpha = 1$, the function is linear, i.e. $w(p) = p$, if $\alpha < 1$, the function overweights small probabilities and underweights big probabilities, and if $\alpha > 1$, it's the contrary effect.

In the PT framework, the concept of rational EV is transformed into what is known as the Subjective Expected Value (SEV). This transformation incorporates both the subjective utility function and the probability distortion function and is expressed as follows:

$$SEV = \sum_i w(p_i).U(x_i)$$

All models are fitted to the monkeys' behavioral dataset to evaluate which best captures their behavior. We conducted the fitting process under two conditions:

- **Combined Fitting** : The models are fitted using both gain and loss trials simultaneously. This approach uses a single set of parameters for both domains.
- **Separate Fitting**: The models are fitted using separate sets of parameters for the gain and loss domains. This involves fitting one probability distortion function for the gain domain and another one for the loss domain, effectively doubling the number of parameters. Models fitted under this condition are denoted by a '+' at the end of their names. For example, P1+ indicates the P1 model with two separate α parameters: α_{gain} and α_{loss}.

4.1.4 Model fit

We developed a user-friendly Python framework to fit various Prospect Theory (PT) models to the behavioral dataset. Our approach models the monkeys' probability of choosing one option based on the difference in subjective expected values between the two options. This is achieved using a standard sigmoid function, as illustrated in Figure 4.5, and expressed mathematically as follows:

$$\mathbf{PT}: \quad P(L_{\text{right}}) = \frac{1}{1 + e^{-\mu.(\Delta_{PT} - x_0)}} \tag{4.1}$$

$$\begin{aligned} &With: \\ &\Delta_{PT} = SEV_{right} - SEV_{left} \\ &\qquad = w(p_{\text{right}}) \cdot U(x_{\text{right}}) - w(p_{\text{left}}) \cdot U(x_{\text{left}}) \end{aligned}$$

- $p_{\text{right}}, p_{\text{left}}$: Probability of getting the non-zero outcome of lotteries right and left, respectively.
- $x_{\text{right}}, x_{\text{left}}$: Non-zero outcomes of lotteries right and left, respectively.

- μ: Parameter that regulates the steepness of the sigmoid curve. Monkeys that have learned the task well will have a larger μ, indicating consistent choice of the option with the highest subjective expected value. Monkeys with smaller μ show less consistent choices.

- x_0: Parameter corresponding to the side bias, with larger absolute values indicating a stronger preference for the corresponding side. When $x_0 > 0$, the monkey exhibits a bias towards the right side, when $x_0 < 0$ the monkey exhibit a bias towards the left side.

- Δ_{PT}: Difference in expected values between the lottery right and the lottery left.

- $w(p)$ and $U(x)$: Probability function and subjective utility functions from the Prospect Theory (PT) framework. All PT models listed above will be tested.

All the PT models were fitted with the monkey data then compared against rational choice models, which use a standard sigmoid function based on the difference in expected values.

$$\textbf{SG}: \quad P(L_{\text{right}}) = \frac{1}{1 + e^{-\mu \cdot (\Delta_{\text{rational}} - x_0)}} \tag{4.2}$$

$$\begin{aligned} &With: \\ \Delta_{rational} &= EV_{right} - EV_{left} \\ &= p_{\text{right}} * x_{\text{right}} - p_{\text{left}} * x_{\text{left}} \end{aligned}$$

The rational models serve as a benchmark, representing decision-making under perfect utility maximization without probability distortion or risk attitude biases.

The models are fitted and evaluated for each monkey using Maximum Likelihood Estimation (MLE), a standard statistical method that involves maximizing the likelihood of the observed data given the model parameters. The log-likelihood function is formulated as follows:

$$\mathcal{L}(O|\theta) = -\sum_{i=1}^{n} [y_i \cdot \log(p(\text{o}_i|\theta)) + (1 - y_i) \cdot \log(1 - p(\text{o}_i|\theta))] \tag{4.3}$$

- θ: a set of PT parameter values. $\theta = \rho, \lambda, \alpha$ if the TK model is used; $\theta = \rho, \lambda, \alpha$ if P1 model is used, $\theta = \rho, \lambda, \alpha, \delta$ if P2 model is used and $\theta = \rho, \lambda, \alpha, \delta$ if GE model is used.

- O: the set of observations under consideration.

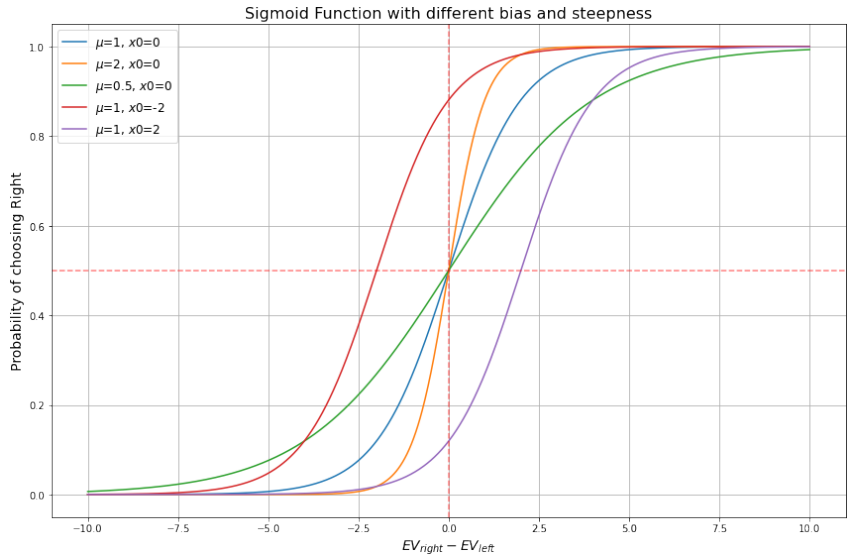

Figure 4.5: The probability of choosing the right option, based on the difference in subjective expected values, is modeled by a sigmoid function. This function includes a parameter μ that regulates the steepness of the sigmoid curve and a parameter $x0$ that represents the bias. Larger values of μ result in a steeper curve, reflecting more consistent decisions. Conversely, smaller values of μ produce a gentler slope, indicating less consistent decision-making. When $x0$ differs from 0, the sigmoid is shifted to the right (if $x0 > 0$) or to the left if ($x0 < 0$), indicating a bias to the left side or the right side, respectively.

- p: the probability according to our decision-making model of making the choice o_i given θ.
- y_i: corresponds to the binary response of the monkey. If $y_i = 1$, the monkey chose the option o_i associated with the probability p, if $y_i = 0$, it did not.

For optimization, the L-BFGS-B (Limited-memory Broyden-Fletcher-Goldfarb-Shanno with Bound constraints) method is employed Zhu et al. [1997]. This algorithm allows to estimate parameters with bound constraints on the variables, by minimizing the negative log-likelihood function. We utilize Scipy library for implementation.

4.1.5 Model evaluation

We evaluate the fit of our models based on three factors:

- **The stability of the models**. Glöckner and Pachur [2012] examined PT models for their parameter stability and predictive accuracy. When comparing the fitting results of two probability distortion functions derived from the

TK expression, they demonstrated that simpler model variants, referring to expressions containing fewer free parameters to be fitted, tend to be more robust and avoiding potential overfitting. Among the models examined, GE and P2 contain the highest number of free parameters, while P1 and TK contain only one free parameter. Consequently, preference should be given to the models with fewer parameters (P1 and TK) due to their potential for better stability and reduced risk of overfitting.

- **The performance of the model to reproduce similar behavior of the monkeys**. We evaluate the fit by comparing the monkeys' responses with the responses generated by the fitted model: we simulate responses from a player based on the fitted model of 1000 randomly chosen trials. For each unique trial, we compute the mean response of the player, we compare it with the mean response. The closer the means are, the better the fit is. The score is then calculated using a custom evaluation metric:

$$\text{Score} = 1 - \frac{1}{N}\sum_{i=1}^{N}|R_{0,i} - R_i|$$

  where:

  - N is the number of unique trials.
  - $R_{0,i}$ is the mean actual response of the monkey for the i-th unique trial.
  - R_i is the mean simulated response of the fitted model for the i-th unique trial.

- **The Bayesian Information Criterion (BIC)** [Vrieze, 2012], which were also used in the work of Tymula et al. [2023]. The BIC is a statistical tool to evaluate the goodness of a fit. It allows to compute accuracy with the fit, while selecting the least complex model to avoid overfitting, favor simple models as it imposed stronger penalty for complex models with too many free parameters:

$$\text{BIC} = n \cdot \log(\text{MSE}) + k \cdot \log(n)$$

  - n is the number of data points,
  - MSE is the mean squared error between the true monkey's responses and the responses generated by the fitted model.
  - k is the number of free parameters in the model,
  - log is the natural logarithm.

4.2 Results

4.2.1 Model selection

Figure 4.6 illustrates the relationship between the number of trials used for fitting and the corresponding fitting error. The analysis revealed that a minimum of 1500 trials is necessary to achieve satisfactory fit precision. Given this threshold, the models are fitted using data from all lottery types introduced in Section 4.1.1. By incorporating all available lottery data, the sample size is significantly increased, leading to more stable and consistent fits across models.

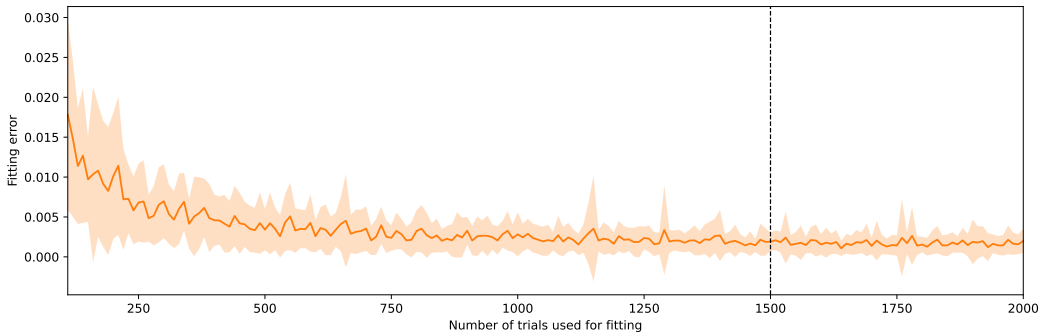

Figure 4.6: **Fit precision according to the number of trials neededl**. 4.1.3

The evaluation metrics were computed for both the rational model (SG) and various PT models (TK, P1, P2, GE, TK+, P2+, GE+). BIC were computed for each model and each monkey. The mean and standard deviation per model are depicted in Figure 4.7. *t*-tests were applied for each mean score against the others to determine whether one mean score was significantly different from the others. Among the PT models, there were no significant differences in their BIC, indicating that they perform similarly well in fitting the data.

Players	SG	TK	P1	P2	GE	TK+	P1+	P2+	GE+
Mean BIC	-92576.0	-95906.0	-96816.0	-96690.0	-97845.0	-97655.0	-97019.0	-99200.0	-99598.0
Std BIC	85110.0	90028.0	90014.0	90874.0	90528.0	90261.0	90098.0	92251.0	92792.0

Figure 4.7: **Mean BIC and std BIC for each model**. The models are fitted for each monkey, this results is the average score of all monkeys. SG: sigmoid model, which shows a rational behavior. TK, P1, P2, GE, TK+, P2+, GE+ are PT models that were introduced in section 4.1.3

The score were computed for all models and shown in Figure 4.8. *t*-tests were also applied for each mean score against the others to determine whether one mean score was significantly different from the others. Notably, the results show that the SG model is significantly worse than the other mean scores (t-statistic = 5.69, p-value <0.001), surpassing the critical T of 2.3646. This suggests that the monkeys

exhibit PT behavior rather than rational behavior, as the PT models outperform the rational model. Furthermore, among the PT models, there were no significant differences in their scores, indicating that they perform similarly well in fitting the data.

Players	SG	TK	P1	P2	GE	TK+	P1+	P2+	GE+
Mean Scores	0.85	0.898	0.913	0.9	0.916	0.906	0.915	0.906	0.926
Std Scores	0.046	0.036	0.03	0.086	0.031	0.033	0.029	0.087	0.031

Figure 4.8: **Mean score and std score of each model**. The models are fitted for each monkey, this results is the average score of all monkeys. SG: sigmoid model, which shows a rational behavior. TK, P1, P2, GE, TK+, P2+, GE+ are PT models that were introduced in section 4.1.3

Considering the equivalent scores among the PT models and the stability criteria introduced in Section 4.1.5, we simplify our analysis by not keeping the most complex models with a high number of parameters. Specifically, P2 and GE (composed of 2 parameters) were not kept as well as their variants P2+ and GE+ (composed of 4 parameters). The P1 and TK models are retained (composed of 1 parameter) to assess the minimum complexity required for capturing the observed behavior. Additionally, the P1+ and TK+ models (composed of 2 parameters) were kept to investigate whether separating gain and loss domains is necessary to effectively capture different behavioral patterns in the data.

4.2.2 Monkey selection

All monkeys were able to do the task properly and with good performances in the control trials. Nevertheless, some monkeys exhibit a strong side bias and were then consequently excluded from the analysis. We define a rejection criteria based on the model scores and the monkeys bias, calculated as follows:

$$bias = abs(n_{right} - n_{left})/n_{tot}$$

- n_{right} number of lottery right chosen.
- n_{left} number of lottery left chosen.
- n_{tot} total number of trials

When a monkey exhibits rational behavior, the SG model (Equation 4.2) is sufficient to explain their decision-making process. This concept is illustrated in Figure 4.9, which depicts the score difference between SG and P1 models in both control and risky trials. In control lotteries, where risk-taking behavior is not involved and

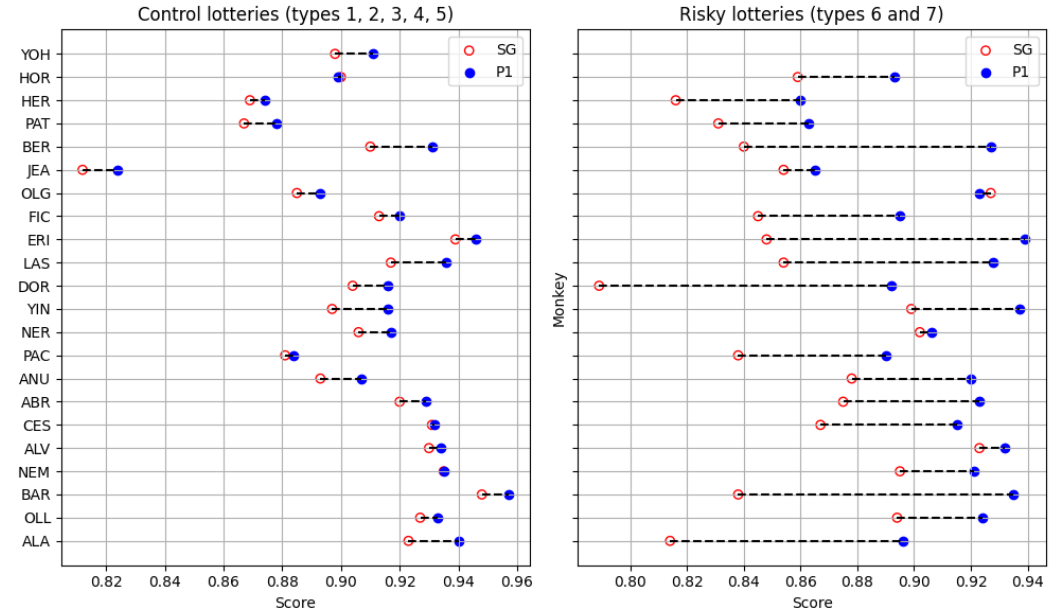

Figure 4.9: **Score difference between SG (rational) and P1 (PT) models in control and risky trials.** The SG model represents the rational choice-making approach, which selects options based on maximizing expected value (Equation 4.2), and the P1 model is grounded in prospect theory, a framework that takes into account psychological biases (Equation 4.1). For each monkey, a segment connecting the two scores reflects the difference $\Delta(SG - P1)$ between the two scores. **Right**: In control lotteries, this difference is small, which is expected because these lotteries do not involve risk-taking behavior; there is always an optimal choice. Therefore, the SG model is sufficient to assess behavior in these trials. **Left**: In risky lotteries, the difference in scores largely increases. This is because the SG model fails to capture the monkeys' behavior in risky situations. In contrast, PT models are better suited to capture this behavior, explaining the substantial increase in score when using these lotteries. This is not the case for highly biased monkeys like NER and OLG for the reasons explained above. The OLA and GAN monkeys from this fitting process because their number of trials shown in Figure 4.3 were below the established threshold of 1500 trials. Notably, similar patterns emerge when comparing the SG model with TK, TK+, and P1+ models.

an optimal choice always exists, the difference between SG and P1 model scores is minimal. This small difference indicates that the SG model adequately captures the monkeys' behavior in these straightforward scenarios. Conversely, in risky lotteries, we observe a substantial increase in the score difference between SG and P1 models for most monkeys. This divergence arises because the SG model fails to accurately represent the monkeys' behavior under risk. PT models, such as P1, suit better to capture this risk-taking behavior, explaining the significant improvement in score. However, this pattern doesn't hold for highly biased monkeys such as NER, OLG

and JEA (their bias values are shown in Figure 4.10). For these subjects, the difference between SG and P1 model scores remains relatively small even in risky trials. This suggests that for monkeys with strong biases, the simpler SG model remains sufficient to explain their behavior, even in risky situations. This phenomenon is illustrated in Figure 4.10-left, where the value of the bias significantly correlates with the SG model score when fitting lottery types 6 and 7.

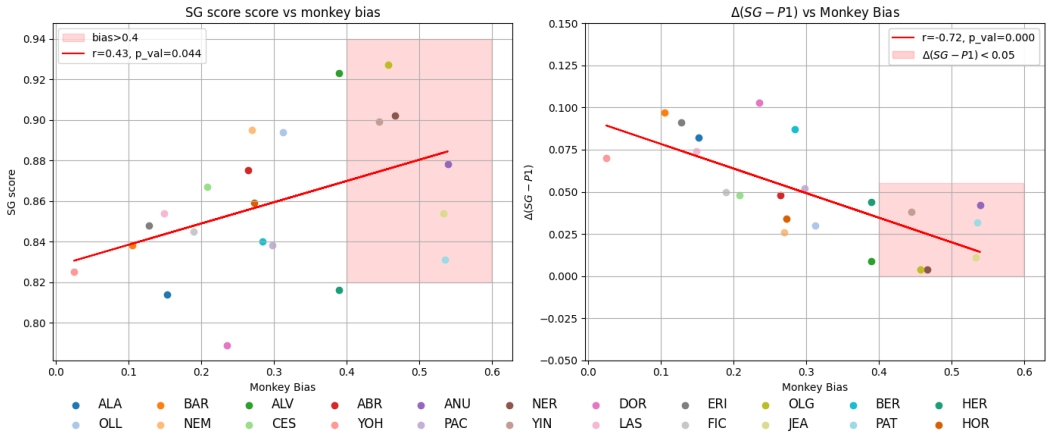

Figure 4.10: **Relationship between monkey bias and model performances when testing on lotteries of type 6/7**. Each color correspond to one monkey. The OLA and GAN monkeys from this fitting process because their number of trials shown in Figure 4.3 were below the established threshold of 1500 trials. **Right**: SG score according to the monkey bias. A significant linear regression indicates that monkeys with higher biases tend to fit better with the SG model, suggesting that the simple rational model is more effective for biased individuals. The red zone indicates rejected models based the criteria $bias > 0.4$. **Left**: The difference in scores between the SG and P1 models correlates significantly with monkey bias, as demonstrated by a linear regression analysis. This relationship indicates that as monkey bias increases, the performance of the SG and P1 models becomes increasingly similar. When the bias has a dominant influence on monkey behavior, the SG model - containing the $x0$ parameter -becomes sufficient to capture the monkey's behavior. In such cases, the additional parameters incorporated in the P1 model may become redundant. The red zone indicates rejected models based the criteria $\Delta(SG - P1) < 0.05$. Notably, similar patterns emerge when comparing the SG model with TK, TK+, and P1+ models.

This correlation indicates that monkeys exhibiting a high bias align well with the SG model in risky situations.This relationship can be attributed to the dominant influence of the side bias on the monkey's behavior. When the side bias is the primary factor driving behavior, the simpler SG model proves sufficient to capture this phenomenon. In such cases, the additional complexity offered by PT models may become unnecessary. While PT models also incorporate the bias parameter $x0$ in the sigmoid function, allowing them to capture side-biased behavior, their

additional parameters related to subjective utility and probability distortion have minimal influence on the monkey's behavior, potentially rendering the PT model fit superfluous. This observation is highlighted in Figure 4.10-left, where the difference in score between SG and P1 ($\Delta(SG - P1)$) significantly decreases as the monkey's bias increases, indicating a diminishing difference in scores between the models as the bias grows.

Notably, similar observations emerge when comparing the SG model with TK, TK+, and P1+ models instead of P1. These observations led us to define two criteria allowing to select the valid monkeys for the final PT analysis:

- $\Delta(SG - P1) > 0.05$
- $bias < 0.4$

The red zone in Figure 4.10 corresponds to the area where the monkeys were rejected. In total, 6 monkeys were rejected: Anubus (ANU), Nereis (NER), Yin, Olga (OLG), and Jeanne (JEA), Patsy (PAT). We keep the 18 other monkeys for the final evaluation of the PT parameters.

4.2.3 Fitting results

The scores of all models are depicted in Figure 4.11. The red values correspond to the worst scores, while the blue values represent the best scores. The monkey 'Yoh' was not well-fitted with the TK model, so it was excluded from the TK model's analysis. All SG fits are shown in red, indicating they are the least effective models for each monkey. This suggests that the monkeys' behavior is better explained by the PT framework than by the rational sigmoid model. Among the models, the TK+ model frequently achieves the highest scores.

Table 4.2 presents the mean values of each model parameter, considering all monkeys (excluding Yoh for the TK analysis). The consistency in mean parameters across all models confirms their stability and suggests that they are capturing similar behavioral patterns.

The models indicate that the monkeys performed the task with high precision, as evidenced by the steep sigmoid slope ($\bar{\mu} = 1.52$) and a slight negative side bias ($\bar{x_0} = -0.31$). On average, all monkeys exhibited:

- Risk aversion in the gain domain ($\bar{\rho_g} = 0.80 < 1$)
- Loss aversion ($\bar{\lambda} = 2.16 > 1$)

However, the parameters of probability distortion vary depending on the model used. Given that the TK and P1 models employ different functions, their parameters

Players	SG	P1	TK	P1+	TK+
ALA	0.891	0.922	0.922	0.925	0.923
OLL	0.878	0.889	0.888	0.890	0.897
BAR	0.901	0.925	0.923	0.924	0.925
NEM	0.880	0.885	0.889	0.887	0.886
ALV	0.900	0.909	0.909	0.911	0.926
CES	0.889	0.900	0.899	0.900	0.910
ABR	0.856	0.864	0.868	0.864	0.869
YOH	0.821	0.842	0.553	0.843	0.851
ANU	0.889	0.909	0.934	0.911	0.936
OLA	0.753	0.769	0.775	0.768	0.776
PAC	0.869	0.887	0.894	0.888	0.902
NER	0.903	0.914	0.916	0.914	0.918
YIN	0.898	0.919	0.922	0.918	0.925
DOR	0.878	0.913	0.923	0.924	0.929
LAS	0.901	0.935	0.936	0.936	0.942
ERI	0.916	0.946	0.949	0.946	0.958
FIC	0.897	0.915	0.921	0.922	0.926
OLG	0.895	0.901	0.903	0.900	0.904
JEA	0.820	0.834	0.833	0.834	0.835
BER	0.892	0.932	0.932	0.931	0.934
PAT	0.851	0.868	0.878	0.869	0.888
HER	0.858	0.874	0.891	0.878	0.885
HOR	0.889	0.899	0.917	0.899	0.904
GAN	0.787	0.794	0.801	0.795	0.807

Figure 4.11: **Result fit of TK model**.

Model Parameters	**P1**	**TK**	**P1+**	**TK+**
μ	1.56	1.49	1.52	1.54
x_0	-0.29	-0.35	-0.30	-0.31
ρ_g	0.80	0.79	0.80	0.82
ρ_l	0.44	0.55	0.45	0.59
λ	2.25	2.06	2.31	2.00
α_g	-	-	0.98	1.15
α_l	-	-	0.59	0.83
α	0.70	1.00	-	-

Table 4.2: Mean values of the PT parameters for all selected monkeys.

are not directly comparable. Nonetheless, comparisons can be made within each model by separating the gain and loss domains. The P1+ and TK+ models reveal that the distortion parameters differ between the gain and loss domains. Specifically, models that separate gains and losses provide distinct parameter values for each domain. In contrast, models that aggregate both gains and losses give parameter

values that fall between those observed in the models that separate the domains. For example, comparing the distortion parameter α in models TK and TK+, we find $\alpha = 1$ for TK, while for TK+ we have $\alpha_g = 1.15$ for gains and $\alpha_l = 0.83$ for losses. The average of these values, $\text{mean}(\alpha_g, \alpha_l) = 0.99$, is almost identical to the overall α in TK+. A similar pattern emerges when comparing models P1 and P1+. The TK and TK+ models are illustrated in Figures 4.13 and 4.14, respectively. While the sigmoid and subjective utility functions remain consistent across both models, the probability distortion function varies depending on whether the trials involve gains or losses. This variation highlights that the monkeys perceive probabilities differently in the gain and loss domains. In the gain domain, the general trend indicates a less distorted perception, with monkeys being generally neutral to probabilities. Conversely, in the loss domain, the global perception appears to be more prone to underestimating probabilities.

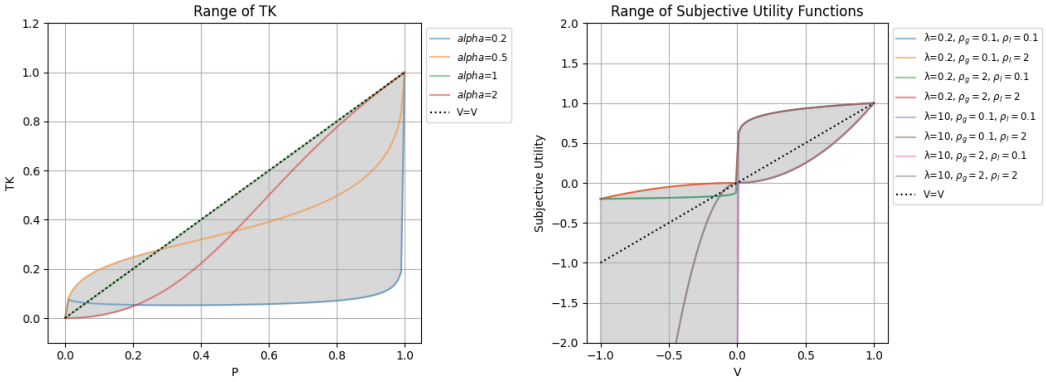

Figure 4.12: **Potential configurations of the TK model. Left:** TK probability distortion function for various α values. **Right:** Subjective utility function for different combinations of λ, ρ_g, and ρ_l. The shaded grey areas represent the range of possible values the model can assume.

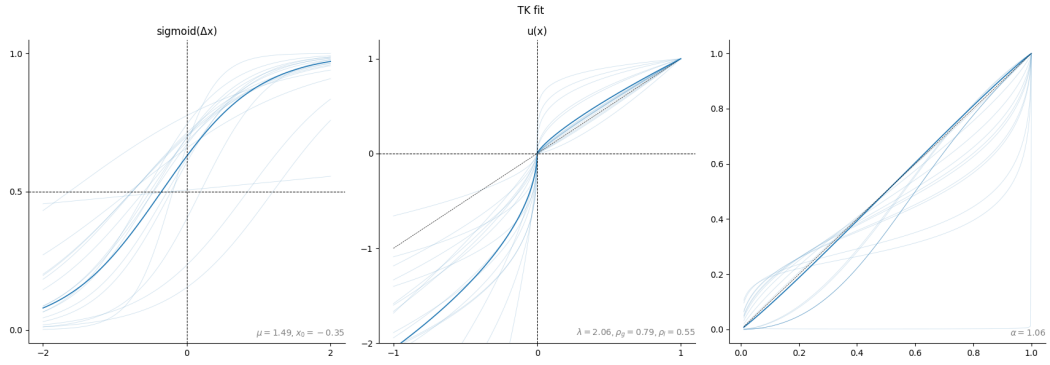

Figure 4.13: **TK Model Fitting Results. From left to right**: fitted sigmoid function, subjective utility function, and probability distortion function. The thick blue line represents the mean fit across all monkeys, while the thin light blue lines depict individual monkey fits. The mean sigmoid is shifted from the right, highlighting a global side bias. The mean utility function is concave in the gain domain ($\rho_g < 1$), indicating global risk-averse behavior. Additionally, an asymmetry between the loss and gain domains is observed ($\lambda < 1$), suggesting a general loss-averse tendency. The mean probability distortion is nearly neutral ($\alpha \approx 1$), indicating that, on average, the individual variability in probability distortion cancels out, resulting in an overall neutral distortion. However, individual monkeys display highly variable distortion functions: some overestimate large probabilities while underestimating small ones, whereas others show the opposite behavior.

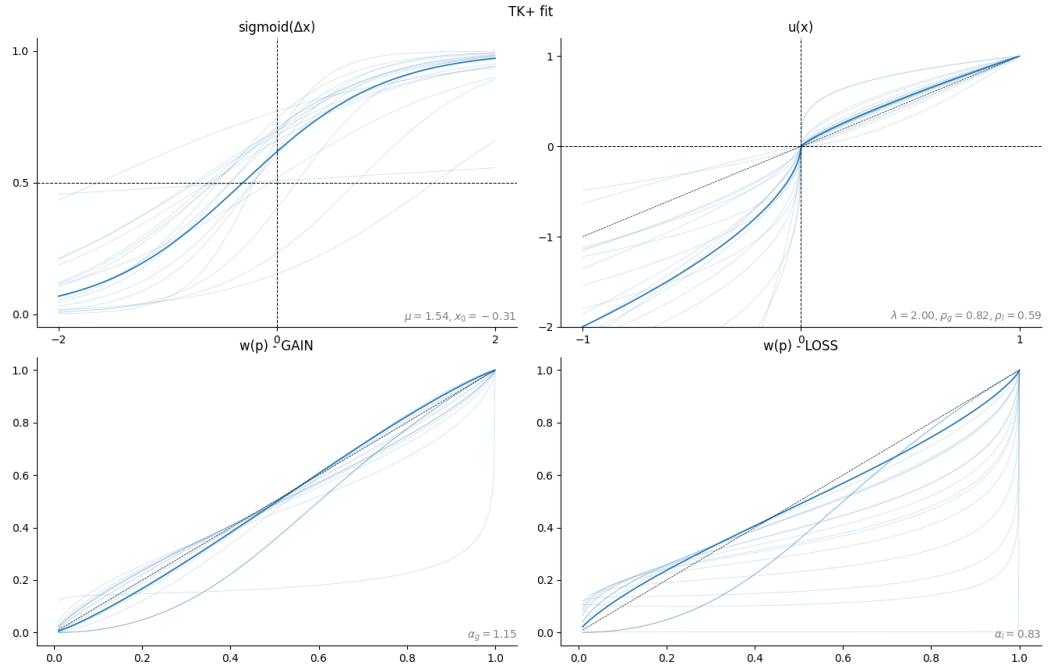

Figure 4.14: **TK+ Model Fitting Results. Top row**: fitted sigmoid function, subjective utility function. **Bottom row**: probability distortion in the gain and in the loss domain. The thick blue line represents the mean fit across all monkeys, while the thin light blue lines depict individual monkey fits. The risk-aversion in the gain domain, loss aversion, side bias are retrieved, similarly to the fitting process without separating gain and loss domain. Nevertheless, an asymmetry is found if the probability distortion: in the gain domain, monkeys overestimate large probabilities while underestimating small ones in average, whereas in the loss domain, it is the opposite effect.

4.2.4 Inter-individual variability

On average, we observe that the monkeys' risk-taking behavior aligns with the PT principles. Specifically, they exhibit risk-averse behavior in the gain domain and loss aversion. Additionally, they demonstrate a distorted perception of probabilities that differs asymmetrically between the gain and loss domains, which contrasts with the typical PT patterns observed in humans. While these findings represent the overall tendencies, Figure 4.14 illustrates significant variability in risk-taking behaviors among individual monkeys. We hypothesize that external factors, particularly the hierarchical structure within monkey groups, may underlie this inter-individual variability and play a crucial role in shaping individual risk-taking behaviors.

Hierarchy in Tonkean macaques

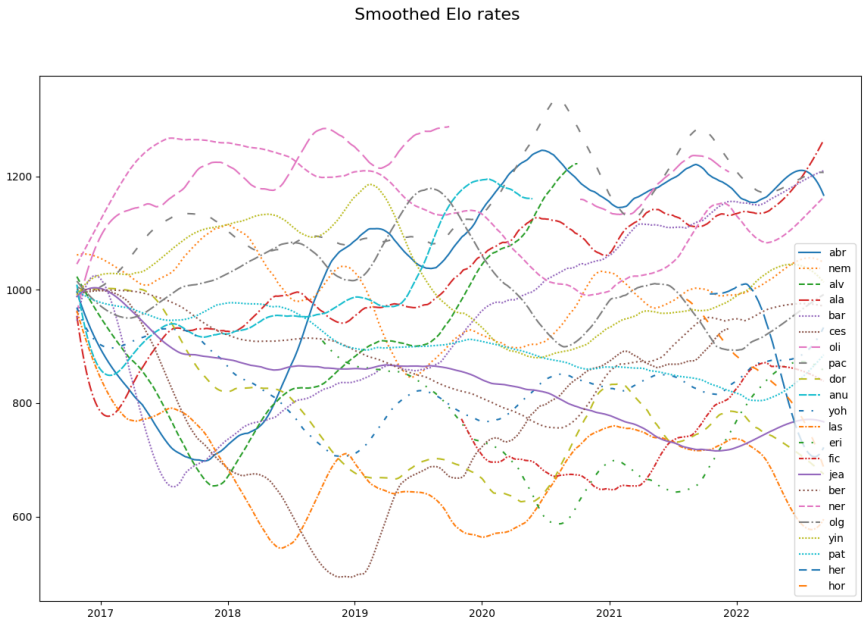

Figure 4.15: **Elo-plot across time from 2017 to end of 2022**. Smoothing has been applied to each line for visibility by using a Savitzky-Golay filter with a window length of 10 and the order of the polynome used to fit the sample is of 3.

Dominance is a crucial social factor that influences interactions among group members in primate communities. Ballesta et al. [2021] introduced a method to measure social relationships and group structure in the Tonkean macaque community using the MALT device. In the semi-free ranging park, monkeys could access the MALT to perform a previously introduced economic task. During this task, one monkey could be replaced by another, reflecting the dominance relationship between them. The monkey that allows itself to be displaced or is forced by another monkey

to leave the MALT is considered subordinate to the monkey taking its place. These displacements indicate social hierarchy and dominance relationships, providing a means to assess the evolution of individual social status within the group. Thus, using MALT allows to measure hierarchy automatically. The Elo-rating is a score used for assessing hierarchy and tracking its evolution across time. Similar to the system used in chess, the higher the Elo-rating, the higher the monkey's position in the hierarchy. Following the method of Ballesta et al. [2021], the evolution of the monkeys' Elo-ratings was computed as depicted in Figure 4.15, showing that each monkey's hierarchy is dynamic and can change over time. For instance, from 2018 to 2022, Abricot (ABR) consistently increased in the hierarchy, whereas Nereis (NER) dropped in 2021 but rose again in 2022. The curves are smoothed using the Savitzky-Golay filter. The Elo-rates can be converted into rankings, with the most dominant monkey (highest elo-score) being ranked first, and the least dominant monkey (lowest Elo-rate) being ranked last, as depicted in Figure 4.16, shown the evolution of ranking between July 2020 to October 2022.

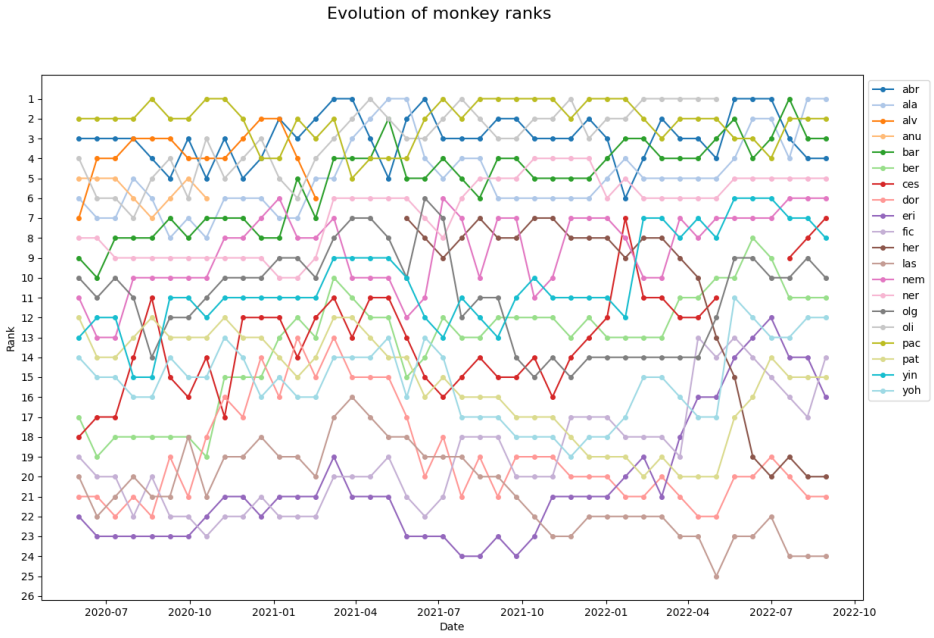

Figure 4.16: **Monkey rank evolution through time from 2020 to 2022** . The most dominant monkey is assigned rank 1, while the most subordinate monkey is given the lowest rank.

Hierarchy vs. risk taking behavior: a static analysis with linear regression

We propose that social hierarchy may play a significant role in shaping risk-taking behavior among monkeys. Specifically, we hypothesize that subordinate and dominant individuals may exhibit distinct behaviors in response to their environment. For instance, subordinate monkeys may experience increased fear due to their lower status, leading them to be more cautious and avoid taking risks. This could manifest in behaviors such as looking back when dominant monkeys approach or playing more cautiously to minimize potential losses. Alternatively, subordinate monkeys may be more risk-averse due to the emotional impact of losing, as their social hierarchy may amplify feelings of fear and anxiety related to loss. As a result, they may be more careful about what they stand to lose and be less likely to take risks. Furthermore, we also consider the possibility that subordinate monkeys have limited access to food resources, making the stakes higher during games involving liquid reward and influencing their behavior. In this scenario, subordinate monkeys may either take more risks in an attempt to secure a larger reward or exhibit reduced risk-taking behavior due to the increased uncertainty of winning.

To verify the relationship between social hierarchy and risk-taking behavior, we first examine the mean Elo-rate of each monkey as a function of their risk-taking parameter, ρ, as shown in Figure 4.17. Notably, we did not detect any significant effect of Elo-score on ρ. Nevertheless, upon closer inspection of both gain and loss domains, we observed an asymmetry that prompted us to separate these domains for further investigation. We hypothesize that the elo-score has a positive impact on the risk-taking parameter ρ and no effect in the loss domain. However, this simple linear regression may not be sufficient as it considers only the elo-score as potential factors that could influence the risk-taking behaviors of monkeys.

To determine whether hierarchy significantly impacts risk-taking behavior, it is crucial to account for all potential influencing factors. We will refine our analysis by incorporating a broader range of variables that may affect risk-taking behavior:

- The social hierarchy
- The age
- The gender
- The probability distortion
- The bias

To increase the reliability of our analysis, we analyzed a 1,500-trial window per monkey during which they performed optimally. This enabled us to compute the

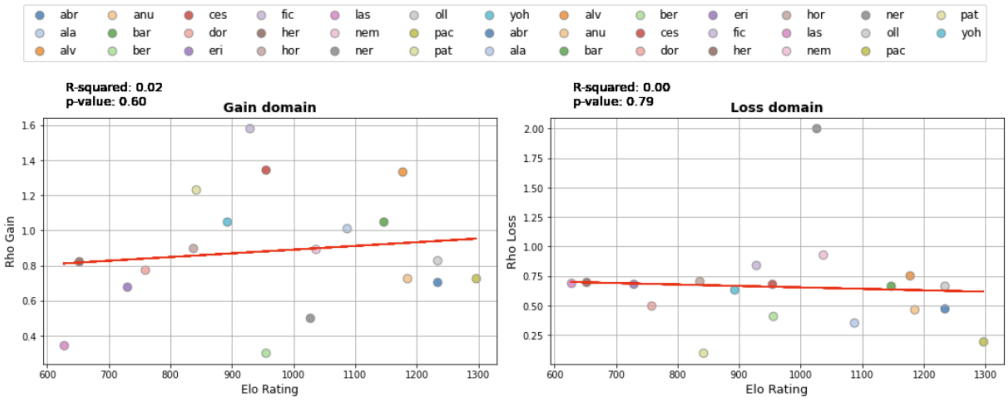

Figure 4.17: **Risk aversion parameter ρ according to the elo-rate.** In the gain domain, dominant monkeys exhibited lower risk-aversion (ρ closer to 1), this suggests that dominant monkeys are less risk-averse when they are in a position of power. However, no significant effect of social hierarchy on risk aversion was observed in the loss domain

corresponding mean Elo-score and PT parameters with greater precision. We employed Linear Mixed Models (LMMs) [2], a statistical method well-suited for analyzing complex data structures. The selection of the best model was performed using the "dredge" function from the "MuMIn" package (version 1.48.4) We examined the effect of the listed factors on both the distortion parameter α and the risk aversion parameter ρ in both gain and loss domains. Although no significant relationships were found with α, our analysis focused on the risk-aversion parameter ρ, which is inversely to monkey risk-aversion, with smaller values indicating greater risk-aversion. The LMM is represented thanks to the following equation:

$$\rho_{+/-} = \beta_0 + \beta_1(\text{elo}) + \beta_2(\text{age}) + \beta_3(\text{sex}) + \beta_4(\alpha) + \beta_5(\text{x0}) + \epsilon. \tag{4.4}$$

Where:

- $\rho_{+/-}$ corresponds to the risk-aversion parameters
- β_0 is the intercept (fixed effect)
- $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the coefficients for the fixed effects: elo-score, age, gender, probability distortion α, and bias $x0$ respectively.

[2] We used LMMs on R with the package lme4. The selection of the best model was performed using the "dredge" function from the MulMIn package, which allows the exploration of different factor combinations to identify the most efficient one. This model selection process is based on quality criteria such as the BIC, which compares models by taking into account both the goodness of fit and the model's complexity, penalizing those with too many parameters. The selected model thus optimizes the balance between prediction accuracy and simplicity

The results of the LMM analysis on ρ_+ are depicted in Figure 4.2.4. Two significant effects were found:

- *elo* has a statistically significant effect on ρ_+ with a p-value $= 0.024 < 0.05$, underlying a positive relationship between hierarchy and risk-taking behavior, as larger values of ρ_+ indicate less risk-aversion.
- α_g has a statistically significant effect on ρ_+ with a p-value $= 0.031 < 0.05$, underlying a negative relationship between distortion and risk-taking behavior.

This correlations raises two possible explanations: firstly, dominant monkeys may be more willing to take risks when facing situations involving gains, potentially suggesting a causal link between dominance and risk-taking behavior. Alternatively, it could be that monkeys that exhibit higher risk-taking tendencies are more likely to ascend to dominant positions. In addition, our findings suggest that the relationship between distortion and risk aversion is complex. Two plausible interpretations emerge: either reduced distortion leads to increased risk aversion or conversely, less risk-aversion results in lower levels of distortion. Notably, our results did not reveal a corresponding effect in the loss domain.

best 1500 TK+

		rho_g	
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.67368	-0.28622 – 1.63359	0.147
elo	0.00117	0.00019 – 0.00214	**0.024**
age	-0.09015	-0.19490 – 0.01460	0.083
sex	0.19028	-0.31126 – 0.69183	0.413
x0	-0.39721	-0.85618 – 0.06175	0.082
alpha g	-0.53540	-1.00958 – -0.06122	**0.031**
alpha 1			
Observations	15		
R^2 / R^2 adjusted	0.741 / 0.597		

Figure 4.18: **LMM analysis: hierarchy and distortion effects on risk-taking behavior**
The LMM analysis, conducted using 1,500 best trials per monkey, reveals statistically significant effects of Elo-score (*elo*) and alpha-group (α_g) on the risk-aversion parameter ρ_+ (Figure 4.2.4). The corresponding p-values are 0.024 and 0.031, respectively, both of which are below the conventional threshold of 0.05. These findings indicate that hierarchy has a positive influence on risk-taking behavior, whereas distortion exhibits a negative effect.

Hierarchy vs. risk taking behavior: a dynamic analysis with linear regression

We subsequently proceeded to a difference-in-difference (DiD) model [Roth et al., 2023]. In this context, the LMM model compares the changes in a risk-taking parameter across different monkeys over time. Rather than selecting a single period per monkey, we divided the dataset into multiple time intervals for each monkey and computed the corresponding mean Elo-score and PT parameters for each specific interval. This approach enabled us to increase the number of observations per monkey, thereby allowing us to examine the dynamic relationships between the parameters. Figure 4.19 provides an example of this period selection process for one individual monkey. The figure displays the evolution of Elo-score over time in blue and the number of trials per day in black lines. The red and blue zones represent low and high Elo-score periods, respectively. By computing PT parameters within these selected periods, we were able to aggregate all results into a comprehensive table, as shown in Figure 4.20.

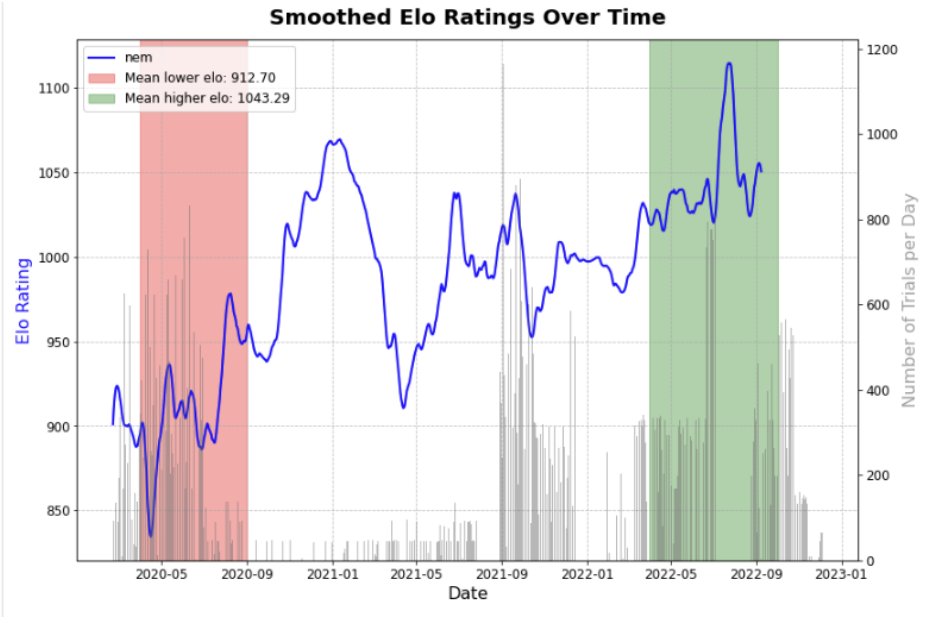

Figure 4.19: **Selection of different periods corresponding to different elo-scores.**

	monkey	elo	date_start	date_end	x0	lambda_	rho_g	rho_l	alpha_g	alpha_l	sex	age
0	abr	1235	10/04/2020	03/10/2020	-0.249303	1.208520	0.409035	0.715509	0.803178	0.452138	1	6.907598
1	abr	1157	19/10/2020	12/05/2021	-1.070870	1.890236	0.516761	0.938080	2.000000	0.306228	1	7.433265
2	abr	1203	11/10/2021	17/12/2021	-0.852618	1.420115	0.624693	0.940909	0.813692	2.000000	1	8.410678
3	ala	1115	11/02/2020	18/08/2020	-0.020719	0.708658	0.767045	0.601515	1.003511	2.000000	1	6.899384
4	ala	1074	12/09/2020	08/03/2021	-0.219892	0.889412	0.835552	0.521146	1.004006	2.000000	1	7.485284
5	ala	1173	28/03/2021	06/06/2021	-0.280463	0.929575	0.629249	0.429996	0.873236	2.000000	1	8.024641
6	ala	1128	02/09/2021	21/03/2022	-0.357240	1.584394	0.932246	0.402032	1.094466	0.652765	1	8.457221
7	ala	1215	07/06/2022	30/08/2022	-0.352368	1.420184	1.031437	0.411749	1.086772	0.602061	1	9.218344
8	alv	1082	04/02/2020	02/06/2020	0.875929	0.441235	0.723077	1.220936	0.843570	0.240705	1	6.976044

Figure 4.20: **Storage of different elo-scores and corresponding PT parameters per monkey and per period. Several row can correspond to the same monkey since we selected several periods for each monkey.**

The LMM incorporates fixed effects (previous listed parameters), but also random effects to account for the inherent variability across individual monkeys, as each monkey was observed over multiple periods. This time, the model accounts for the difference in parameters instead of the parameters themselves. This approach allows us to model the differences in parameters rather than the parameters themselves. The LMM enables us to assess whether differences in specific independent variables (i.e. Elo-score) are associated with differences in the dependent variable (i.e. risk-taking behavior). Mathematically, this can be represented by the following equation:

$$\Delta(\rho_{+/-}) = \beta_0 + \beta_1 \Delta(\text{elo-score}) + \beta_2 \Delta(\text{age}) + \beta_3 \Delta(\text{sex}) + \beta_4 \Delta(\alpha) + \beta_5 \Delta(x0) + u_{(\text{monkey id})} + \epsilon \tag{4.5}$$

Where:

- $\Delta(\rho_{+/-})$ corresponds to the difference between two risk-aversion parameters of two different periods.
- β_0 is the intercept (fixed effect)
- $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the coefficients for the difference of the fixed effects: $\Delta(\text{elo-score}), \Delta(\text{age}), \Delta(\text{sex}), \Delta(\alpha), \Delta(x0)$ respectively.
- $u_{(\text{monkey id})}$ is the random effect associated with the monkey identity.

dynamic TK+

	delta_rho_g		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.10164	-0.80536 – 0.60209	0.751
delta elo	0.00141	0.00004 – 0.00277	**0.045**
delta age	0.38189	0.04544 – 0.71834	**0.030**
sex	-0.23120	-0.70406 – 0.24166	0.297
delta x0	0.06772	-0.16323 – 0.29868	0.524
delta alpha g	0.37486	-0.10042 – 0.85014	0.108
Observations	15		
R^2 / R^2 adjusted	0.661 / 0.472		

Figure 4.21: **LMM analysis: hierarchy and age effects on risk-taking behavior** The LMM analysis, conducted using multiple periods per monkey, reveals statistically significant effects of Elo-score (*elo*) and age on the risk-aversion parameter ρ_+. The corresponding p-values are 0.045 and 0.030, respectively, both of which are below the conventional threshold of 0.05. These findings indicate that increase in hierarchy has a positive influence on risk-taking behavior, and increase in age also exhibits a positive effect.

The results from our DiD LMM model are displayed in Figure 4.2.4. Two significant effects were found:

- $\Delta(age)$ has a statistically significant effect on $\Delta(\rho_+)$ with a p-value $= 0.030 < 0.05$, underlying a significant positive association between age and risk-taking behavior in the gain domain.

- $\Delta(elo - score)$ has a statistically significant positive impact on $\Delta(\rho_+)$, with a p-value of $0.045 < 0.05$, underlying a significant positive association between hierarchy and risk-taking behavior

Hierarchy does affect monkey's behavior

These results suggest a potential causal link between dominance/age and risk-taking behavior in the gain domain (notably, these results did not reveal a corresponding effect in the loss domain). Since larger values of ρ_+ indicate less risk-aversion, these findings suggest that as monkeys get older, they become more inclined to take risks when facing situations involving gains. This result implies that the aging process is associated with decreased risk-aversion. Nevertheless, further verification are needed to figure out the direction of causality in the case of social hierarchy: does the dominant status lead to more risk-taking attitude, or is it because monkeys take risks that they end up to a better hierarchical position ? Overall, the results align with our hypothesis stating that social hierarchy may have an effect on the risk taking behavior. The place of the monkey in the hierarchy induce conditions in which it does the economic task by modifying its behavior.

Furthermore, our investigation reveals that these studied parameters are not static entities, but rather dynamic and susceptible to change throughout the life of the monkeys. The social hierarchy of the monkeys is indeed a dynamic factor that evolves over time, as well as the parameters associated with risk-taking behavior. Specifically, we find that changes in social ranking can induce shifts in mental states, ultimately influencing an individual monkey's risk-taking behavior. For instance, monkeys that ascend to higher ranks through repeated social victories tend to become more confident and less risk-averse over time. Conversely, those that decline in rank may lose confidence and become more risk-averse.

4.3 Discussion

We developed a user-friendly tool to analyze risk-taking behavior and distortion of probability perception in our large dataset. We analyzed the risk attitudes of 24 macaque monkeys of varying ages and genders using an economic task designed to present risky situations. Our findings indicate that, overall, the monkeys exhibit risk aversion when faced with potential gains and risk-seeking behavior when confronted with potential losses. Additionally, they demonstrated an inverted S-shaped probability distortion pattern. These results confirm that monkey behavior aligns with the PT framework. Since the latter comprises different models in the literature, our tool enabled us to test all of them, compare them, and define a precise method to determine which one is the best. We built additional models allowing separation of gain and loss domains in the task, which confirmed an asymmetry of treatment between the two domains [Nioche et al., 2019, 2021].

Furthermore, we defined a precise method to select monkeys for the PT analysis. Some monkeys showed a very strong side bias (e.g., Yin, Nereis, Anubis, Patricia selecting the same cue >75% of the time; Olga and Jeanne >90% of the time), and we hypothesized that that they may have a tendency towards maintaining the status quo [Samuelson and Zeckhauser, 1988]. Status quo bias is a cognitive bias where individuals favor a default option, even when alternatives might be more advantageous [Kahneman et al., 1991, Samuelson and Zeckhauser, 1988]. This aligns with research indicating that nonhuman primates, like humans, can exhibit status quo bias [Englund, 2023]. The status quo is often associated with loss aversion [Kahneman et al., 1991], where the monkeys might perceive switching sides as a potential "loss" compared to the "gain" of sticking with their habitual choice. However, attributing this side bias to status quo bias should be done with caution, as our experimental setup did not explicitly define a clear default, i.e. 'status quo' option. Side biases in primates could also arise from other factors such as handedness [Deuel and Schaffer, 1987] or spatial preferences.

Finally, while we observed a general trend conforming to the PT framework, we also discovered significant inter-individual variability among the monkeys. We suggested that this variability might arise from environmental factors such as social dynamics. To investigate this hypothesis, we employed a novel method for tracking social hierarchy over time and obtained the Elo-score of each monkey at various points in time. By computing the corresponding PT parameters for these periods and using LMM we found that the social hierarchy exhibits a significant positive correlation with the risk-aversion of the monkeys in the gain domain. Specifically, our results demonstrate that dominant monkeys tend to be less risk-averse. This finding was further supported by observing that the social hierarchy is not static, but rather evolves over time, and that this evolution is significantly correlated with the evolution of the monkey's risk-aversion. Our study provides strong evidence for a potential link between social status and risk attitudes in macaque monkeys,

suggesting that social factors can influence their behavior. Notably, these results were not found in the loss domain, revealing an asymmetry in the relationship between risk-taking behavior and social hierarchy, with a significant effect in the gain domain, but not in the loss domain. The underlying reasons for this asymmetry remain unclear and need further investigation. Nevertheless, we believe that our framework provides a promising starting point for exploring this unanswered question. Furthermore, we identified additional parameters, including age and probability distortion, which significantly impact the risk-aversion parameter. While we did not pursue further analysis of these findings at this stage, they will be an essential focus for future research.

5 Conclusion

This thesis explores the relationship between brain network structures and the cognitive process of decision-making, addressing a fundamental question in neuroscience: how does the architecture of neural networks influence decision-making?

The study of brain connectivity is a critical area of research, as underscored in the introduction. It provides deep insights into the functional implications of brain structure in both healthy and dysfunctional brains. Advances in neuroimaging, neuro-electrophisology and computational modeling have allowed to reveal some important structural properties of the brain, including random and non random properties, hierarchical and modular structures. It has also been shown that some brain regions have been conserved throughout evolution, suggesting that they are crucial for the maintenance of fundamental cognitive processes.

However, the underlying reasons behind these structural properties and their persistence across evolution remain opened. This thesis aims to contribute to this discussion. At the modeling level, we addressed two questions: is a specific network structure necessary to solve any given decision task? Could alternative architectures emerge to achieve similar functional and neuronal properties?

Throughout this thesis, we adopted an approach that minimizes strong assumptions about network architecture. This was made possible by employing the reservoir computing paradigm, which offered several key advantages. First, it incorporates biological plausibility through the use of recurrent neural networks with random connectivity, mirroring the brain's inherent recurrence and randomness. Second, it allowed us to avoid overly restrictive architectural assumptions. Third, its simplicity in design, manipulation, and training enabled us to explore a wide range of structures and training variations, allowing us to test and train a diverse set of models.

In the first phase of our research, we developed a minimal reservoir model to investigate decision-making processes in a spatial navigation context. This model consisted of a pool of randomly interconnected neurons tasked with solving a T-maze alternation problem. The task was implemented in a continuous state-space, significantly increasing its complexity. The model received only partially observable information, further enhancing the challenge and more closely mimicking real-world scenarios. The readout model was trained using supervised learning. While

effective, this approach lacks biological plausibility. To address this limitation, a still ongoing project aims to replace this training method with RL. Several studies have already explored the integration of reservoir computing with RL in continuous state-spaces: Szita et al. [2006] provided a theoretical framework demonstrating the viability of this approach. Schmidt et al. [2014], Oubbati et al. [2014], Dasgupta et al. [2013] implemented actor-critic architectures where the reservoir is the critic, outputting action values. Koprinkova-Hristova [2014], Koprinkova-Hristova et al. [2013], Koprinkova-Hristova [2015, 2016] introduced various approaches including adaptive, heuristic, and Hebbian learning mechanisms in the context of reservoir computing and RL. Implementing RL and reservoir computing in a continuous state-space presents significant challenges. However, this approach is crucial for more accurately modeling real-world cognitive processes, and will be one of the major perspective of this work.
Despite its basic structure and lack of specific structural constraints, the model performed the T-maze task well. This success indicates that, for this particular task, no specific structure beyond the constraint of recurrence is necessary. Additionally, upon analyzing the neural properties within our model, we identified neurons exhibiting characteristics akin to hippocampal splitter cells, which are known to encode an animal's past experiences and trajectory. This finding challenges the notion that certain neural properties - here splitter cells - are exclusive to specific brain structures - here the hippocampus - and, most importantly, suggests that splitter cell-like activity may be more of a consequence or artifact of navigational behavior rather than its cause, challenging the traditional interpretation of these neural properties of the hippocampus.

The success of our simple model underscores the potential of a minimalistic approach and paved the way of the second phase of our research, which explored whether more complex behaviors could also emerge from this minimal model. To explore this, we introduced a new task based on the 2-arm bandit paradigm, which the minimal model solved with high performance. We subsequently added complexity to the task, by adding a binding problem and random temporal constraints. The performance of the minimal model significantly decreased. We introduced more structural constraints, rendering the structure more forward-oriented either by building deep reservoir models or by changing directly the input connectivity of one reservoir. We implemented fast and slow pathways, similar to those in the BG. The structured models demonstrated significantly better performance on the complex task compared to the unstructured models. Nevertheless, the structured models were only trained on one specific type of task which constitute our main limitation in this research. Our perspective will be to test these architectures on other decision-making tasks involving temporal and speed constraints such as the temporal order judgement (TOJ) [Hendrich et al., 2012] in order to verify whether these results are generalized to other tasks. While our results are not generalized, they suggest that as task difficulty increases, more structured networks become

necessary to maintain high performance.

In the third phase of our research, we explored whether these required structures emerged as unique optimal solutions. We introduced a third task within the context of economic decision-making, which enabled us to assess both the performance of our models in decision-making and their risk-taking behavior. We implemented the NEAT algorithm to evolve the structure of the reservoir models based on evolutionary principles, which allowed us to see the emergence of at least to distinct types of structural solutions to solve the task, demonstrating that, in our set up, multiple realizability is shown. Additionally, we studied the risk-taking behavior of these two types of structures emerging from the population and showed that at the individual level, different structure type can lead to the same risk-taking behavior, answering the question of multiple realizability. However, significant differences at the population level indicated that the structure of decision-making networks can substantially influence risk-taking behavior. These findings led us to conclude that in our context, structural variations affect decision-making behavior at the population level.
Evolving networks structures with NEAT was highly time-consuming and required HPC to process the evolution. To resolve this issue, our future direction for this project is to utilize HyperNEAT [Stanley et al., 2009], an advanced method that allows for the evolution of reservoir structures through indirect encoding of neural networks. This approach is not only faster but also facilitates the generation of larger neural networks. By employing HyperNEAT, we will explore different set of parameters and conditions, including the evolution of hyperparameters that remained fixed in our current study.

One of the main outcome of the NEAT process was the emergence of a balance between risk-seeking and risk-averse behaviors. This result closely mirrors observations in humans and other species, suggesting that certain behavioral patterns are vital for evolutionary success. This finding has important implications and connections to existing research, as it may help explain the persistence of individual differences in risk attitudes within populations. It prompted us to investigate how this balance manifests and how it might be influenced by environmental factors. As our final project, we shifted our focus from structural questions to a broader inquiry: if structural factors influence behavior at the population level, but individual behaviors within the same species vary, what other factors might drive these behavioral differences? Thanks to our extensive behavioral data collection from 24 monkeys, we were able to investigate risk-taking behavior both at the population level and at the individual level. Our analysis revealed that social factors, including the hierarchical structure within the monkey community, have a significant impact on risk-taking behavior. Specifically, we found that dominant monkeys tend to be less risk-averse, but also that monkeys who experience social victories become less risk-averse over time, while those who lose social status over time become more risk-averse. This dynamic aspect highlights the relationship

between social hierarchy and risk-taking behavior. However, it is essential to note that the direction of causality remains unclear: do dominant monkeys exhibit less risk aversion thanks to their social position, or does their confidence leading them to take greater risks, that ultimately put them into a dominant status? To shed more light on this question, a more precise analysis at a finer time scale would be beneficial. For instance, observing how individual monkeys adjust their behavior before and after a social victory or loss could provide valuable insights into the direct impact of these events on confidence and risk-taking. Additionally, exploring other factors that may influence risk-taking behavior is crucial. Our preliminary findings suggest that age, distortion probability also play a role in shaping risk aversion. Reaction times were also shown as preliminary results, as we believe that they could bring more information about the monkeys' behavior. Future research directions should focus on delving deeper into the analysis of these parameters and conducting more precise analyses over time to better understand the dynamic relationships.

Among the monkey population, we identified some individuals who were notably biased. Due to time constraints, we could not study their behavior in greater detail. Future research could delve deeper into these behaviors, potentially confirming or refuting our hypothesis about status quo bias and further exploring the heuristics involved in economic decision-making.

This thesis is an attempt to adopt an approach that progresses from minimal to more complex models, reflecting the evolutionary trajectory of biological neural systems. This approach aligned closely with Cisek [2019]'s perspective on studying theories of behavior. Cisek argues that understanding brain function necessitates an evolutionary perspective [Cisek and Hayden, 2022], advocating for a method called "phylogenetic refinement." His approach aims to build theories based on the actual evolutionary history of organisms and move away from traditional psychological distinctions. Indeed, this evolutionary perspective contrasts with top-down approaches that start with complex human behaviors and artificially separate cognitive functions. Phylogenetic refinement instead, starts with simple mechanisms and elaborating them to more complex ones, ensuring that proposed mechanisms are compatible with known evolutionary changes, with each new mechanism conceived as an extension of an ancestral one.
Additionally to this introduced evolutionary aspect, this thesis introduced the concept of degeneracy in various reservoir computing networks. This concept aligned with the work of Onur Güntürkün on avian cognition, which has demonstrated principles of multiple realizability and degeneracy in bird brains. Güntürkün et al. [2021] show that despite structural differences from mammalian brains, avian brains have independently developed similar cognitive abilities and mechanisms to support them, demonstrating that different structural configurations can yield similar functional outcomes. This finding suggests that despite distinct evolutionary origins, complex cognition can evolve through different neural architectures, indicating mul-

tiple pathways to achieving intelligence [Güntürkün, 2012, 2005]. Similarly, Farris [2008] demonstrated that insect mushroom bodies exhibit structural, functional, and developmental convergence with the brain of vertebrates, a phenomenon likely due to the adaptation to common behavioral ecologies across these diverse animal groups.

We believe that coupling evolutionary modeling with the analysis of experimental behavioral data is presented as a promising approach. Developing artificial networks that are adaptable to varying environmental conditions and based on evolutionary processes, would help identifying common computational principles that emerge across different neural architectures, and could help elucidate the impact of ecological constraints in shaping neural organization. Additionally, insights gained from studying the evolution of ANN could not only inform neuroscience research but also lead to the discovery of efficient, robust, and adaptable AI architectures capable of solving complex cognitive tasks through different structural approaches.

Bibliography

Eric Aislan Antonelo and Benjamin Schrauwen. On learning navigation behaviors for small mobile robots with reservoir computing architectures. *IEEE Transactions on Neural Networks and Learning Systems*, 26(4):763–780, Apr 2015. ISSN 2162-2388. doi: 10.1109/tnnls.2014.2323247. URL `http://dx.doi.org/10.1109/TNNLS.2014.2323247`.

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631, 2019.

Roger L Albin, Anne B Young, and John B Penney. The functional anatomy of basal ganglia disorders. *Trends in neurosciences*, 12(10):366–375, 1989.

Garrett E Alexander, Mahlon R DeLong, and Peter L Strick. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*, 9(1):357–381, 1986.

Eric Antonelo and Benjamin Schrauwen. Learning slow features with reservoir computing for biologically-inspired robot localization. *Neural Networks*, 25: 178–190, Jan 2012. ISSN 0893-6080. doi: 10.1016/j.neunet.2011.08.004. URL `http://dx.doi.org/10.1016/j.neunet.2011.08.004`.

Eric A Antonelo and Benjamin Schrauwen. Unsupervised learning in reservoir computing: Modeling hippocampal place cells for small mobile robots. In *Artificial Neural Networks–ICANN 2009: 19th International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings, Part I 19*, pages 747–756. Springer, 2009.

Eric A Antonelo, Benjamin Schrauwen, and Jan Van Campenhout. Generative modeling of autonomous robots and their environments using reservoir computing. *Neural Processing Letters*, 26:233–249, 2007.

Eric A Antonelo, Benjamin Schrauwen, and Dirk Stroobandt. Event detection and localization for small mobile robots using reservoir computing. *Neural Networks*, 21(6):862–871, 2008.

Dirk Jan Ardesch, Lianne H Scholtens, and Martijn P Van Den Heuvel. The human connectome from an evolutionary perspective. *Progress in brain research*, 250: 129–151, 2019.

Adam R Aron and Russell A Poldrack. Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *Journal of Neuroscience*, 26(9), 2006.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

Gökhan Aydogan, Remi Daviet, Richard Karlsson Linnér, Todd A Hare, Joseph W Kable, Henry R Kranzler, Reagan R Wetherill, Christian C Ruff, Philipp D Koellinger, BIG BEAR Consortium, et al. Genetic underpinnings of risky behaviour relate to altered neuroanatomy. *Nature Human Behaviour*, 5(6):787–794, 2021.

Ke Bai, Fangzhou Liao, and Xiaolin Hu. Reservoir computing with a small-world network for discriminating two sequential stimuli. In *Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21–26, 2017, Proceedings, Part I 14*. Springer, 2017.

Bernard W Balleine, Mauricio R Delgado, and Okihide Hikosaka. The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience*, 27(31): 8161–8165, 2007.

Sébastien Ballesta, Baptiste Sadoughi, Fabia Miss, Jamie Whitehouse, Géraud Aguenounon, and Hélène Meunier. Assessing the reliability of an automated method for measuring dominance hierarchy in non-human primates. *Primates*, 62:595–607, 2021.

Izhar Bar-Gad, Genela Morris, and Hagai Bergman. Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Progress in neurobiology*, 71(6), 2003.

Omri Barak. Recurrent neural networks as versatile tools of neuroscience research. *Current opinion in neurobiology*, 46:1–6, 2017.

Pablo Barttfeld, Bruno Wicker, Sebastián Cukier, Silvana Navarta, Sergio Lew, Ramón Leiguarda, and Mariano Sigman. State-dependent changes of connectivity patterns and functional brain network topology in autism spectrum disorder. *Neuropsychologia*, 50(14):3653–3662, 2012.

Danielle Smith Bassett and ED Bullmore. Small-world brain networks. *The neuroscientist*, 12(6), 2006.

Hannah M Bayer and Paul W Glimcher. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47(1):129–141, 2005.

Alison M Bell. Approaching the genomics of risk-taking behavior. *Advances in genetics*, 68:83–104, 2009.

Daniel Bennett, Yael Niv, and Angela J Langdon. Value-free reinforcement learning: policy optimization as a minimal model of operant behavior. *Current Opinion in Behavioral Sciences*, 41:114–121, 2021.

James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 24, 2011.

Rafal Bogacz and Kevin Gurney. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural computation*, 19(2):442–477, 2007.

Thomas Boraud. *L'Matière à décision*. CNRS éditions, 2015.

Thomas Boraud. *How the Brain Makes Decisions*. Oxford University Press, USA, 2020.

Thomas Boraud, Arthur Leblois, and Nicolas P Rougier. A natural history of skills. *Progress in Neurobiology*, 171:114–124, 2018.

Valentino Braitenberg. *Vehicles: Experiments in synthetic psychology*. MIT press, 1986.

Signe Brinkløv, M Brock Fenton, and John M Ratcliffe. Echolocation in oilbirds and swiftlets. *Frontiers in physiology*, 4:123, 2013.

Scott D. Brown and Andrew Heathcote. The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57, 11 2008. ISSN 00100285. doi: 10.1016/j.cogpsych.2007.12.002.

Adrian Bruhin, Helga Fehr-Duda, and Thomas Epper. Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78(4):1375–1412, 2010.

György Buzsáki and Edvard I Moser. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature neuroscience*, 16(2):130–138, 2013.

Thomas Caraco. Energy budgets, risk and foraging preferences in dark-eyed juncos (junco hyemalis). *Behavioral Ecology and Sociobiology*, 8:213–217, 1981.

Andrea Ceni, Peter Ford Dominey, Claudio Gallicchio, Alessio Micheli, Luca Pedrelli, and Domenico Tortorella. Continuously deep recurrent neural networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 59–73. Springer, 2024.

Naomi Chaix-Echel, Snigdha Dagar, Frédéric Alexandre, Thomas Boraud, and Nicolas P Rougier. On the epistemic role of hippocampal cells: the case of splitter cells. *bioRxiv*, pages 2024–06, 2024.

Naomi Chaix-Eichel, Gautham Venugopal, Thomas Boraud, and Nicolas P Rougier. Exploring the role of structure in a time constrained decision task. *arXiv preprint arXiv:2401.10849*, 2024.

V Srinivasa Chakravarthy and Ahmed A Moustafa. *Computational neuroscience models of the basal ganglia*. Springer, 2018.

Kyriakos C Chatzidimitriou and Pericles A Mitkas. A neat way for evolving echo state networks. In *ECAI 2010*, pages 909–914. IOS Press, 2010.

Mario Chavez, Miguel Valencia, Vincent Navarro, Vito Latora, and Jacques Martinerie. Functional modularity of background activities in normal and epileptic brain networks. *Physical review letters*, 104(11):118701, 2010.

Zhenbo Cheng, Zhidong Deng, Xiaolin Hu, Bo Zhang, and Tianming Yang. Efficient reinforcement learning of a reservoir network model of parametric working memory achieved with a cluster population winner-take-all readout mechanism. *Journal of Neurophysiology*, 114(6), 2015.

Mark M. Churchland and Krishna V. Shenoy. Preparatory activity and the expansive null-space. *Nature Reviews Neuroscience*, 25(4):213–236, March 2024. ISSN 1471-0048. doi: 10.1038/s41583-024-00796-z. URL `http://dx.doi.org/10.1038/s41583-024-00796-z`.

Paul Cisek. Resynthesizing behavior through phylogenetic refinement. *Attention, Perception, & Psychophysics*, 81:2265–2287, 2019.

Paul Cisek and Benjamin Y Hayden. Neuroscience needs evolution, 2022.

Paul Cisek, Geneviève Aude Puskas, and Stephany El-Murr. Decisions in changing conditions: The urgency-gating model. *Journal of Neuroscience*, 29, 2009. ISSN 02706474. doi: 10.1523/JNEUROSCI.1844-09.2009.

Anne Gabrielle Eva Collins. Reinforcement learning: bringing together computation and cognition. *Current Opinion in Behavioral Sciences*, 29:63–68, 2019.

Anne GE Collins and Michael J Frank. Opponent actor learning (opal): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological review*, 121(3):337, 2014.

Anne GE Collins and Amitai Shenhav. Advances in modeling learning and decision-making in neuroscience. *Neuropsychopharmacology*, 47(1):104–118, 2022.

Christine M Constantinople, Alex T Piet, and Carlos D Brody. An analysis of decision under risk in rats. *Current Biology*, 29(12):2066–2074, 2019.

José L Contreras-Vidal and Wolfram Schultz. A predictive reinforcement model of dopamine neurons for learning approach behavior. *Journal of computational neuroscience*, 6:191–214, 1999.

Steven J Cook, Travis A Jarrell, Christopher A Brittin, Yi Wang, Adam E Bloniarz, Maksim A Yakovlev, Ken CQ Nguyen, Leo T-H Tang, Emily A Bayer, Janet S Duerr, et al. Whole-animal connectomes of both caenorhabditis elegans sexes. *Nature*, 571(7763):63–71, 2019.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

Matthew Dale, Simon O'Keefe, Angelika Sebald, Susan Stepney, and Martin A Trefzer. Reservoir computing quality: connectivity and topology. *Natural Computing*, 20, 2021.

Sakyasingha Dasgupta, Florentin Wörgötter, Jun Morimoto, and Poramate Manoonpong. Neural combinatorial learning of goal-directed behavior with reservoir critic and reward modulated hebbian plasticity. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pages 993–1000. IEEE, 2013.

William Dauer and Serge Przedborski. Parkinson's disease: mechanisms and models. *Neuron*, 39(6):889–909, 2003.

Jacob D. Davidson and Ahmed El Hady. Foraging as an evidence accumulation process. *PLOS Computational Biology*, 15(7), 07 2019. doi: 10.1371/journal.pcbi.1007060.

Peter Dayan. Improving generalization for temporal difference learning: The successor representation. *Neural computation*, 5(4):613–624, 1993.

Zhidong Deng and Yi Zhang. Collective behavior of a small-world recurrent neural system with scale-free distribution. *IEEE Transactions on neural networks*, 18(5), 2007.

Ruthmary K Deuel and Scott P Schaffer. Patterns of hand preference in monkeys. *Behavioral and Brain Sciences*, 10(2):270–271, 1987.

Yacila I Deza Araujo, Stephan Nebe, Philipp T Neukam, Shakoor Pooseh, Miriam Sebold, Maria Garbusow, Andreas Heinz, and Michael N Smolka. Risk seeking for losses modulates the functional connectivity of the default mode and left frontoparietal networks in young males. *Cognitive, Affective, & Behavioral Neuroscience*, 18:536–549, 2018.

Peter Ford Dominey. A connectivity gradient in structured reservoir computing predicts a hierarchy for mixed selectivity in human cortex. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2024.

Peter Ford Dominey, Timothy M Ellmore, and Jocelyne Ventre-Dominey. Effects of connectivity on narrative temporal processing in structured reservoir computing. In *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022.

Éléonore Duvelle, Roddy M Grieves, and Matthijs AA Van der Meer. Temporal context and latent state inference in the hippocampal splitter signal. *ELife*, 12: e82357, 2023.

Gerald M Edelman and Joseph A Gally. Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, 98(24):13763–13768, 2001.

Victor M Eguiluz, Dante R Chialvo, Guillermo A Cecchi, Marwan Baliki, and A Vania Apkarian. Scale-free brain functional networks. *Physical review letters*, 94(1): 018102, 2005.

Agoston E Eiben and James E Smith. *Introduction to evolutionary computing*. Springer, 2015.

Lotem Elber-Dorozko and Yonatan Loewenstein. Striatal action-value neurons reconsidered. *ELife*, 7:e34248, 2018.

Pierre Enel, Emmanuel Procyk, René Quilodran, and Peter Ford Dominey. Reservoir computing properties of neural dynamics in prefrontal cortex. *PLoS computational biology*, 12(6):e1004967, 2016.

Maisy D Englund. An exploration of the status quo bias in nonhuman primates. 2023.

Umberto Esposito, Michele Giugliano, Mark Van Rossum, and Eleni Vasilaki. Measuring symmetry, asymmetry and randomness in neural network connectivity. *PloS one*, 9(7):e100805, 2014.

WE Evans. Echolocation by marine delphinids and one species of fresh-water dolphin. *The Journal of the Acoustical Society of America*, 54(1):191–199, 1973.

Shiva Farashahi, Habiba Azab, Benjamin Hayden, and Alireza Soltani. On the flexibility of basic risk attitudes in monkeys. *Journal of Neuroscience*, 38(18): 4383–4398, 2018.

Sarah M Farris. Structural, functional and developmental convergence of the insect mushroom bodies with higher brain centers of vertebrates. *Brain Behavior and Evolution*, 72(1):1–15, 2008.

Simone Ferrari-Toniolo, Philipe M Bujold, and Wolfram Schultz. Probability distortion depends on choice sequence in rhesus monkeys. *Journal of Neuroscience*, 39(15):2915–2929, 2019.

Thomas HB FitzGerald, Philipp Schwartenbeck, and Raymond J Dolan. Reward-related activity in ventral striatum is action contingent and modulated by behavioral relevance. *Journal of Neuroscience*, 34(4):1271–1279, 2014.

Jonas Fizet, Adam Rimele, Thierry Pebayle, Jean-Christophe Cassel, Christian Kelche, and Hélène Meunier. An autonomous, automated and mobile device to concurrently assess several cognitive functions in group-living non-human primates. *Neurobiology of learning and memory*, 145:45–58, 2017.

Anders M Fjell and Kristine B Walhovd. Structural brain changes in aging: courses, causes and cognitive consequences. *Reviews in the Neurosciences*, 21(3):187–222, 2010.

Alex Fornito, Jong Yoon, Andrew Zalesky, Edward T Bullmore, and Cameron S Carter. General and specific functional connectivity disturbances in first-episode schizophrenia during cognitive control performance. *Biological psychiatry*, 70(1): 64–72, 2011.

Loren M Frank, Emery N Brown, and Matthew Wilson. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, 27(1):169–178, 2000.

Michael J Frank. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural networks*, 19(8):1120–1136, 2006.

Claudio Gallicchio and Alessio Micheli. Echo state property of deep reservoir computing networks. *Cognitive Computation*, 9:337–350, 2017.

Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, 268, 2017.

Dileep George, Rajeev V Rikhye, Nishad Gothoskar, J Swaroop Guntupalli, Antoine Dedieu, and Miguel Lázaro-Gredilla. Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nature communications*, 12(1):2392, 2021.

Samuel J Gershman and Yael Niv. Learning latent structure: carving nature at its joints. *Current Opinion in Neurobiology*, 20(2):251–256, April 2010. ISSN 0959-4388. doi: 10.1016/j.conb.2010.02.008. URL `http://dx.doi.org/10.1016/j.conb.2010.02.008`.

Benoît Girard, Nicolas Tabareau, Quang-Cuong Pham, Alain Berthoz, and J-J Slotine. Where neuroscience and dynamic system theory meet autonomous robotics: a contracting basal ganglia model for action selection. *Neural Networks*, 21(4): 628–641, 2008.

Andreas Glöckner and Thorsten Pachur. Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, 123(1):21–32, 2012.

William M Goldstein and Hillel J Einhorn. Expression theory and the preference reversal phenomena. *Psychological review*, 94(2):236, 1987.

Matthew D Golub, Patrick T Sadtler, Emily R Oby, Kristin M Quick, Stephen I Ryu, Elizabeth C Tyler-Kabara, Aaron P Batista, Steven M Chase, and Byron M Yu. Learning by neural reassociation. *Nature neuroscience*, 21(4):607–616, 2018.

Sten Grillner and Brita Robertson. The basal ganglia over 500 million years. *Current Biology*, 26(20):R1088–R1100, 2016.

Sten Grillner, Brita Robertson, and Marcus Stephenson-Jones. The evolutionary origin of the vertebrate basal ganglia and its role in action selection. *The Journal of physiology*, 591(22):5425–5431, 2013.

Onur Güntürkün. Avian and mammalian "prefrontal cortices": limited degrees of freedom in the evolution of the neural mechanisms of goal-state maintenance. *Brain research bulletin*, 66(4-6):311–316, 2005.

Onur Güntürkün. The convergent evolution of neural substrates for cognition. *Psychological research*, 76(2):212–219, 2012.

Onur Güntürkün, Kaya von Eugen, Julian Packheiser, and Roland Pusch. Avian pallial circuits and cognition: a comparison to mammals. *Current Opinion in Neurobiology*, 71:29–36, 2021.

Kevin Gurney, Tony J Prescott, and Peter Redgrave. A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological cybernetics*, 84:401–410, 2001.

Martin Guthrie, Arthur Leblois, André Garenne, and Thomas Boraud. Interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study. *Journal of neurophysiology*, 109(12):3025–3040, 2013.

MD György Buzsáki. *The brain from inside out*. Oxford University Press, 2019.

Aric Hagberg and Drew Conway. Networkx: Network analysis with python. *URL: https://networkx. github. io*, 2020.

Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7):e159, 2008.

Lawrence D Harder and Leslie A Real. Why are bumble bees risk averse? *Ecology*, 68(4):1104–1108, 1987.

Tom Hartley, Colin Lever, Neil Burgess, and John O'Keefe. Space in the brain: how the hippocampal formation supports spatial cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1635):20120510, 2014.

David Heeger et al. Poisson model of spike generation. *Handout, University of Standford*, 5(1-13):76, 2000.

Richard P Heitz. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in neuroscience*, 8:86875, 2014.

Elisabeth Hendrich, Tilo Strobach, Martin Buss, Hermann J Müller, and Torsten Schubert. Temporal-order judgment of visual and auditory stimuli: modulations in situations with and without stimulus discrimination. *Frontiers in Integrative Neuroscience*, 6:63, 2012.

Claus C Hilgetag and Alexandros Goulas. Is the brain really a small-world network? *Brain Structure and Function*, 221:2361–2366, 2016.

Xavier Hinaut and Nathan Trouvain. Which hype for my new task? hints and random search for reservoir computing hyperparameters. In *ICANN 2021-30th International Conference on Artificial Neural Networks*, 2021.

Xavier Hinaut, Florian Lance, Colas Droin, Maxime Petit, Gregoire Pointeau, and Peter Ford Dominey. Corticostriatal response selection in sentence production: Insights from neural network simulation with reservoir computing. *Brain and language*, 150:54–68, 2015.

Arend Hintze, Randal S Olson, Christoph Adami, and Ralph Hertwig. Risk sensitivity as an evolutionary adaptation. *Scientific reports*, 5(1):8242, 2015.

Binyamin Hochner. An embodied view of octopus neurobiology. *Current Biology*, 22(20):R887–R892, 2012.

Shogo Homma and Masanori Takezawa. Risk preference as an outcome of evolutionarily adaptive learning mechanisms: An evolutionary simulation under diverse risky environments. *PloS one*, 19(8):e0307991, 2024.

James C Houk, James L Adams, and Andrew G Barto. A model of how the basal ganglia generate and use neural signals that predict reinforcement. 1994.

Marc W. Howard and Michael J. Kahana. A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3):269–299, June 2002. ISSN 0022-2496. doi: 10.1006/jmps.2001.1388. URL `http://dx.doi.org/10.1006/jmps.2001.1388`.

Mark D Humphries, Robert D Stewart, and Kevin N Gurney. A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. *Journal of Neuroscience*, 26(50):12921–12942, 2006.

Makoto Ito and Kenji Doya. Parallel representation of value-based and finite state-based strategies in the ventral and dorsal striatum. *PLoS computational biology*, 11(11):e1004540, 2015.

Herbert Jaeger. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148(34):13, 2001.

Herbert Jaeger. Echo state network. *scholarpedia*, 2(9), 2007.

Sarah Jarvis, Stefan Rotter, and Ulrich Egert. Extending stability through hierarchical clusters in echo state networks. *Frontiers in neuroinformatics*, 4, 2010.

Daphna Joel, Yael Niv, and Eytan Ruppin. Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural networks*, 15(4-6):535–547, 2002.

Anika K Josef, David Richter, Gregory R Samanez-Larkin, Gert G Wagner, Ralph Hertwig, and Rui Mata. Stability and change in risk-taking propensity across the adult life span. *Journal of personality and social psychology*, 111(3):430, 2016.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):363–391, 1979.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5(1):193–206, 1991.

Richard Karlsson Linnér, Pietro Biroli, Edward Kong, S Fleur W Meddens, Robbee Wedow, Mark Alan Fontana, Maël Lebreton, Stephen P Tino, Abdel Abdellaoui, Anke R Hammerschlag, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature genetics*, 51(2):245–257, 2019.

Yuji Kawai, Jihoon Park, and Minoru Asada. A small-world topology enhances the echo state property and signal propagation in reservoir computing. *Neural Networks*, 112, 2019.

Mehdi Khamassi, Loïc Lachèze, Benoît Girard, Alain Berthoz, and Agnès Guillot. Actor-critic models of reinforcement learning in the basal ganglia.

Mehdi Khamassi, Loïc Lachèze, Benoît Girard, Alain Berthoz, and Agnès Guillot. Actor–critic models of reinforcement learning in the basal ganglia: from natural to artificial rats. *Adaptive Behavior*, 13(2):131–148, 2005.

Mehdi Khamassi, René Quilodran, Pierre Enel, Peter Dominey, and Emmanuel Procyk. Role of the frontal cortex in solving the exploration-exploitation trade-off. In *Cinquième conférence plénière française de Neurosciences Computationnelles," Neurocomp'10"*, 2010.

Ken-ichi Kitayama. Guiding principle of reservoir computing based on "small-world" network. *Scientific reports*, 12(1), 2022.

Oren Kolodny and Caitlin Stern. Evolution of risk preference is determined by reproduction dynamics, life history, and population size. *Scientific reports*, 7(1): 9364, 2017.

Vijay Konda and John Tsitsiklis. Actor-critic algorithms. *Advances in neural information processing systems*, 12, 1999.

Petia Koprinkova-Hristova. Adaptive critic design and heuristic search for optimization. In *Large-Scale Scientific Computing: 9th International Conference, LSSC 2013, Sozopol, Bulgaria, June 3-7, 2013. Revised Selected Papers 9*, pages 248–255. Springer, 2014.

Petia Koprinkova-Hristova. Hebbian versus gradient training of esn actors in closed-loop acd. In *Numerical Methods and Applications: 8th International Conference, NMA 2014, Borovets, Bulgaria, August 20-24, 2014, Revised Selected Papers 8*, pages 95–102. Springer, 2015.

Petia Koprinkova-Hristova. Three approaches to train echo state network actors of adaptive critic design. In *Artificial Neural Networks and Machine Learning–ICANN 2016: 25th International Conference on Artificial Neural Networks, Barcelona, Spain, September 6-9, 2016, Proceedings, Part I 25*, pages 494–501. Springer, 2016.

Petia Koprinkova-Hristova, Mohamed Oubbati, and Günther Palm. Heuristic dynamic programming using echo state network as online trainable adaptive critic. *International Journal of Adaptive Control and Signal Processing*, 27(10):902–914, 2013.

Robert A Kozol, Andrew J Conith, Anders Yuiska, Alexia Cree-Newman, Bernadeth Tolentino, Kasey Benesh, Alexandra Paz, Evan Lloyd, Johanna E Kowalko, Alex C Keene, et al. A brain-wide analysis maps structural evolution to distinct anatomical module. *Elife*, 12:e80777, 2023.

Roger Kurlan. Tourette's syndrome: current concepts. *Neurology*, 39(12):1625–1625, 1989.

Francesco Lamanna, Francisca Hervas-Sotomayor, A Phillip Oel, David Jandzik, Daniel Sobrido-Cameán, Megan L Martik, Stephen A Green, Thoomke Brüning, Katharina Mößinger, Julia Schmidt, et al. Reconstructing the ancestral vertebrate brain using a lamprey neural cell type atlas. *BioRxiv*, pages 2022–02, 2022.

Rosamund F Langston, James A Ainge, Jonathan J Couey, Cathrin B Canto, Tale L Bjerknes, Menno P Witter, Edvard I Moser, and May-Britt Moser. Development of the spatial representation system in the rat. *Science*, 328(5985):1576–1580, 2010.

Arthur Leblois, Thomas Boraud, Wassilios Meissner, Hagai Bergman, and David Hansel. Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia. *Journal of Neuroscience*, 26(13):3567–3583, 2006.

Corentin Léger, Gautier Hamon, Eleni Nisioti, Xavier Hinaut, and Clément Moulin-Frier. Evolving reservoirs for meta reinforcement learning. In *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*, pages 36–60. Springer, 2024.

Jian Li and Nathaniel D Daw. Signals in human striatum are appropriate for policy update rather than value prediction. *Journal of Neuroscience*, 31(14):5504–5511, 2011.

Wei Liao, Zhiqiang Zhang, Zhengyong Pan, Dante Mantini, Jurong Ding, Xujun Duan, Cheng Luo, Guangming Lu, and Huafu Chen. Altered functional connectivity and small-world in mesial temporal lobe epilepsy. *PloS one*, 5(1):e8525, 2010.

Xuhong Liao, Athanasios V Vasilakos, and Yong He. Small-world human brain networks: perspectives and challenges. *Neuroscience & Biobehavioral Reviews*, 77:286–300, 2017.

Timothy P Lillicrap, Adam Santoro, Luke Marris, Colin J Akerman, and Geoffrey Hinton. Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6): 335–346, 2020.

Yong Liu, Meng Liang, Yuan Zhou, Yong He, Yihui Hao, Ming Song, Chunshui Yu, Haihong Liu, Zhening Liu, and Tianzi Jiang. Disrupted small-world networks in schizophrenia. *Brain*, 131(4):945–961, 2008.

Mantas Lukoševičius. A practical guide to applying echo state networks. In *Neural Networks: Tricks of the Trade: Second Edition*, pages 659–686. Springer, 2012.

Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer science review*, 3(3):127–149, 2009.

EA Maguire, Richard SJ Frackowiak, and Christopher D Frith. Learning to find your way: a role for the human hippocampal formation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1377):1745–1750, 1996.

Tiago V Maia and Michael J Frank. From reinforcement learning models to psychiatric and neurological disorders. *Nature neuroscience*, 14(2):154–162, 2011.

Luca Manneschi, Matthew OA Ellis, Guido Gigante, Andrew C Lin, Paolo Del Giudice, and Eleni Vasilaki. Exploiting multiple timescales in hierarchical echo state networks. *Frontiers in Applied Mathematics and Statistics*, 6:616658, 2021a.

Luca Manneschi, Andrew C Lin, and Eleni Vasilaki. Sparce: improved learning of reservoir computing systems through sparse representations. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2):824–838, 2021b.

Filip Matzner. Neuroevolution on the edge of chaos. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 465–472, 2017.

Rose McDermott, James H Fowler, and Oleg Smirnov. On the evolutionary origin of prospect theory preferences. *The Journal of Politics*, 70(2):335–350, 2008.

David Meunier, Renaud Lambiotte, and Edward T Bullmore. Modular and hierarchically modular organization of brain networks. *Frontiers in neuroscience*, 4: 7572, 2010.

Sifis Micheloyannis, Ellie Pachou, Cornelis Jan Stam, Michael Breakspear, Panagiotis Bitsios, Michael Vourkas, Sophia Erimaki, and Michael Zervakis. Small-world networks and disturbed functional connectivity in schizophrenia. *Schizophrenia research*, 87(1-3):60–66, 2006.

Jonathan W Mink. The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology*, 50(4):381–425, 1996.

Seyedali Mirjalili and Seyedali Mirjalili. Genetic algorithm. *Evolutionary algorithms and neural networks: Theory and applications*, pages 43–55, 2019.

John Moon, Yuting Wu, and Wei D Lu. Hierarchical architectures in reservoir computing systems. *Neuromorphic Computing and Engineering*, 1(1), 2021.

Edvard I Moser, Emilio Kropff, and May-Britt Moser. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.*, 31:69–89, 2008.

Edvard I Moser, May-Britt Moser, and Bruce L McNaughton. Spatial representation in the hippocampal formation: a history. *Nature neuroscience*, 20(11):1448–1464, 2017.

May-Britt Moser, David C Rowland, and Edvard I Moser. Place cells, grid cells, and memory. *Cold Spring Harbor perspectives in biology*, 7(2):a021808, 2015.

L Muessig, J Hauser, TJ Wills, and F Cacucci. Place cell networks in pre-weanling rats show associative memory properties from the onset of exploratory behavior. *Cerebral Cortex*, 26(8):3627–3636, 2016.

Lyle Muller, Alain Destexhe, and Michelle Rudolph-Lilith. Brain networks: small-worlds, after all? *New Journal of Physics*, 16(10):105004, 2014.

Atsushi Nambu, Hironobu Tokuno, and Masahiko Takada. Functional significance of the cortico–subthalamo–pallidal 'hyperdirect'pathway. *Neuroscience research*, 43(2):111–117, 2002.

Aurélien Nioche, Sacha Bourgeois-Gironde, and Thomas Boraud. An asymmetry of treatment between lotteries involving gains and losses in rhesus monkeys. *Scientific Reports*, 9(1):10441, 2019.

Aurélien Nioche, Nicolas P Rougier, Marc Deffains, Sacha Bourgeois-Gironde, Sébastien Ballesta, and Thomas Boraud. The adaptive value of probability distortion and risk-seeking in macaques' decision-making. *Philosophical Transactions of the Royal Society B*, 376(1819):20190668, 2021.

John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 1971.

Lisa Ordonez and Lehman Benson III. Decisions under time pressure: How time constraint affects risky decision making. *Organizational Behavior and Human Decision Processes*, 71(2):121–140, 1997.

Srdjan Ostojic and Stefano Fusi. Computational role of structure in neural activity and connectivity. *Trends in Cognitive Sciences*, 2024.

Mohamed Oubbati, Christian Fischer, and Günther Palm. Intrinsically motivated decision making for situated, goal-driven agents. In *International Conference on Simulation of Adaptive Behavior*, pages 166–175. Springer, 2014.

R Pascanu. On the difficulty of training recurrent neural networks. *arXiv preprint arXiv:1211.5063*, 2013.

Razvan Pascanu and Herbert Jaeger. A neurodynamical model for working memory. *Neural networks*, 24(2):199–207, 2011.

Benjamin Pasquereau, Agnes Nadjar, David Arkadir, Erwan Bezard, Michel Goillandeau, Bernard Bioulac, Christian Eric Gross, and Thomas Boraud. Shaping of motor responses by incentive values through the basal ganglia. *Journal of Neuroscience*, 27(5):1176–1183, 2007.

Juan Pérez-Fernández, Marta Barandela, and Cecilia Jiménez-López. The dopaminergic control of movement-evolutionary considerations. *International journal of molecular sciences*, 22(20):11284, 2021.

Lucia Perez Rodero. Characterizing brain connectivity during driving risk situations by transfer entropy measures of eeg.

M. Andrea Pisauro, Elsa Fouragnan, Chris Retzler, and Marios G. Philiastides. Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous eeg-fmri. *Nature Communications*, 8, 6 2017. ISSN 2041-1723. doi: 10.1038/ncomms15808.

Drazen Prelec. The probability weighting function. *Econometrica*, pages 497–527, 1998.

Hilary Putnam. Psychological predicates. *Art, Mind, and Religion*, 1:37–48, 1967.

Rajkumar Vasudeva Raju, J Swaroop Guntupalli, Guangyao Zhou, Carter Wendelken, Miguel Lázaro-Gredilla, and Dileep George. Space is a latent sequence: A theory of the hippocampus. *Science Advances*, 10(31):eadm8470, 2024.

Roger Ratcliff and Gail McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20, 4 2008. ISSN 0899-7667. doi: 10.1162/neco.2008.12-06-420.

Nathaniel Rodriguez, Eduardo Izquierdo, and Yong-Yeol Ahn. Optimal modularity and memory capacity of neural reservoirs. *Network Neuroscience*, 3(2), 2019.

Matthew R Roesch, Teghpal Singh, P Leon Brown, Sylvina E Mullins, and Geoffrey Schoenbaum. Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *Journal of Neuroscience*, 29(42):13365–13376, 2009.

Christopher A Ross and Sarah J Tabrizi. Huntington's disease: from molecular pathogenesis to clinical treatment. *The Lancet Neurology*, 10(1):83–98, 2011.

Jonathan Roth, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023.

Nicolas P. Rougier. A density-driven method for the placement of biological cells over two-dimensional manifolds. *Frontiers in Neuroinformatics*, 12, 3 2018. ISSN 1662-5196. doi: 10.3389/fninf.2018.00012.

Javier Salas-Rodríguez, Luis Gómez-Jacinto, Isabel Hombrados-Mendieta, and Natalia del Pino-Brunet. Applying an evolutionary approach of risk-taking behaviors in adolescents. *Frontiers in psychology*, 12:694134, 2022.

Kazuyuki Samejima, Yasumasa Ueda, Kenji Doya, and Minoru Kimura. Representation of action-specific reward values in the striatum. *Science*, 310(5752): 1337–1340, 2005.

William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of risk and uncertainty*, 1:7–59, 1988.

Honi Sanders, Matthew A Wilson, and Samuel J Gershman. Hippocampal remapping as hidden state inference. *Elife*, 9:e51140, 2020.

Philipp Schlegel, Yijie Yin, Alexander S Bates, Sven Dorkenwald, Katharina Eichler, Paul Brooks, Daniel S Han, Marina Gkantia, Marcia Dos Santos, Eva J Munnelly, et al. Whole-brain annotation and multi-connectome cell typing of drosophila. *Nature*, 634(8032):139–152, 2024.

Nico M Schmidt, Matthias Baumgartner, and Rolf Pfeifer. Actor-critic design using echo state networks in a simulated quadruped robot. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2224–2229. IEEE, 2014.

Robert Schmidt, Daniel K Leventhal, Nicolas Mallet, Fujun Chen, and Joshua D Berke. Canceling actions involves a race between basal ganglia pathways. *Nature neuroscience*, 16(8), 2013.

Wolfram Schultz, Peter Dayan, and P Read Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.

Stephen M Smith, Diego Vidaurre, Christian F Beckmann, Matthew F Glasser, Mark Jenkinson, Karla L Miller, Thomas E Nichols, Emma C Robinson, Gholamreza Salimi-Khorshidi, Mark W Woolrich, et al. Functional connectomics from resting-state fmri. *Trends in cognitive sciences*, 17(12):666–682, 2013.

Olaf Sporns. The non-random brain: efficiency, economy, and complex dynamics. *Frontiers in computational neuroscience*, 5:5, 2011.

Olaf Sporns. *Networks of the Brain*. MIT press, 2016.

Olaf Sporns and Richard F Betzel. Modular brain networks. *Annual review of psychology*, 67:613–640, 2016.

Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS computational biology*, 1(4):e42, 2005.

Kimberly L Stachenfeld, Matthew M Botvinick, and Samuel J Gershman. The hippocampus as a predictive map. *Nature neuroscience*, 20(11):1643–1653, 2017.

Cornelis J Stam, BF Jones, Guido Nolte, Michael Breakspear, and Ph Scheltens. Small-world networks and functional connectivity in alzheimer's disease. *Cerebral cortex*, 17(1):92–99, 2007.

Kenneth O Stanley and Risto Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary computation*, 10(2):99–127, 2002.

Kenneth O Stanley, David B D'Ambrosio, and Jason Gauci. A hypercube-based encoding for evolving large-scale neural networks. *Artificial life*, 15(2):185–212, 2009.

Laurence Steinberg. A social neuroscience perspective on adolescent risk-taking. In *Biosocial theories of crime*, pages 435–463. Routledge, 2017.

Kilian Stenning, Jack Gartside, Luca Manneschi, Christopher Cheung, Tony Chen, Jake Love, Alexander Vanstone, Holly Holder, Francesco Caravelli, Karin Everschor-Sitte, et al. Deep physical reservoir computing with programmable

nanomagnetic hierarchies. In *APS March Meeting Abstracts*, volume 2023, pages Q57–014, 2023.

Marcus Stephenson-Jones, Ebba Samuelsson, Jesper Ericsson, Brita Robertson, and Sten Grillner. Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Current biology*, 21(13):1081–1091, 2011.

Michael D Stern. Patrimony and the evolution of risk-taking. *PloS one*, 5(7):e11656, 2010.

Takuma Sumi, Hideaki Yamamoto, Yuichi Katori, Koki Ito, Satoshi Moriya, Tomohiro Konno, Shigeo Sato, and Ayumi Hirano-Iwata. Biological neurons act as generalization filters in reservoir computing. *Proceedings of the National Academy of Sciences*, 120(25):e2217008120, 2023.

Roland E Suri, J Bargas, and MA Arbib. Modeling functions of striatal dopamine modulation in learning and planning. *Neuroscience*, 103(1):65–85, 2001.

Shreyas M Suryanarayana, Juan Pérez-Fernández, Brita Robertson, and Sten Grillner. The lamprey forebrain–evolutionary implications. *Brain Behavior and Evolution*, 96(4-6):318–333, 2022.

David Sussillo. Neural circuits as computational dynamical systems. *Current opinion in neurobiology*, 25:156–163, 2014.

Richard S Sutton and Andrew G Barto. The reinforcement learning problem. *Reinforcement learning: An introduction*, pages 51–85, 1998.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

István Szita, Viktor Gyenes, and András Lőrincz. Reinforcement learning with echo state networks. In *International Conference on Artificial Neural Networks*, pages 830–839. Springer, 2006.

Sumi Takuma, Yamamoto Hideaki, Katori Yuichi, Ito Koki, Sato Shigeo, and Hirano-Iwata Ayumi. Modular topology enhances reservoir computing performance in biological neuronal networks. *IEICE Proceedings Series*, 76(D2L-11), 2023.

Martina Testori, Hedwig Eisenbarth, and Rebecca B Hoyle. Selfish risk-seeking can provide an evolutionary advantage in a conditional public goods game. *Plos one*, 17(1):e0261340, 2022.

David Thura, Julie Beauregard-Racine, Charles William Fradet, and Paul Cisek. Decision making by urgency gating: Theory and experimental support. *Journal of Neurophysiology*, 108, 2012. ISSN 00223077. doi: 10.1152/jn.01071.2011.

Meropi Topalidou, Daisuke Kase, Thomas Boraud, and Nicolas P Rougier. A computational model of dual competition between the basal ganglia and the cortex. *eneuro*, 5(6), 2018.

Rüdiger M Trimpop. *The psychology of risk taking behavior*. Elsevier, 1994.

Nathan Trouvain, Luca Pedrelli, Thanh Trung Dinh, and Xavier Hinaut. ReservoirPy: An efficient and user-friendly library to design echo state networks. In *Artificial Neural Networks and Machine Learning – ICANN 2020*, pages 494–505. Springer International Publishing, 2020a. doi: 10.1007/978-3-030-61616-8_40. URL `https://doi.org/10.1007/978-3-030-61616-8_40`.

Nathan Trouvain, Luca Pedrelli, Thanh Trung Dinh, and Xavier Hinaut. Reservoirpy: an efficient and user-friendly library to design echo state networks. In *International Conference on Artificial Neural Networks*. Springer, 2020b.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5:297–323, 1992.

Agnieszka Tymula, Xueting Wang, Yuri Imaizumi, Takashi Kawai, Jun Kunimatsu, Masayuki Matsumoto, and Hiroshi Yamada. Dynamic prospect theory: Two core decision theories coexist in the gambling behavior of monkeys and humans. *Science Advances*, 9(20):eade7972, 2023.

Marius Usher and James L. McClelland. The time course of perceptual choice: The leaky, competing accumulator model, 2001. ISSN 0033295X.

Martijn P Van den Heuvel, Edward T Bullmore, and Olaf Sporns. Comparative connectomics. *Trends in cognitive sciences*, 20(5):345–361, 2016.

David C Van Essen, Kamil Ugurbil, Edward Auerbach, Deanna Barch, Timothy EJ Behrens, Richard Bucholz, Acer Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

Carl Van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.

Alexandre Variengien and Xavier Hinaut. A journey in esn and lstm visualisations on a language task. *arXiv preprint arXiv:2012.01748*, 2020.

George Velentzas, Costas Tzafestas, and Mehdi Khamassi. Bio-inspired meta-learning for active exploration during non-stationary multi-armed bandit tasks. In *2017 Intelligent Systems Conference (IntelliSys)*, pages 661–669. IEEE, 2017.

David Verstraeten, Benjamin Schrauwen, and Dirk Stroobandt. Reservoir-based techniques for speech recognition. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 1050–1053. IEEE, 2006.

John Von Neumann and Oskar Morgenstern. Theory of games and economic behavior, 2nd rev. 1947.

Scott I Vrieze. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2):228, 2012.

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8:279–292, 1992.

Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

Jeffrey J Wenstrup and Christine V Portfors. Neural processing of target distance by echolocating bats: functional roles of the auditory midbrain. *Neuroscience & Biobehavioral Reviews*, 35(10):2073–2083, 2011.

James CR Whittington, Timothy H Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy EJ Behrens. The tolman-eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183(5):1249–1263, 2020.

Thomas J Wills, Laurenz Muessig, and Francesca Cacucci. The development of spatial behaviour and the hippocampal neural representation of space. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1635):20130409, 2014.

Emma R Wood, Paul A Dudchenko, R Jonathan Robitsek, and Howard Eichenbaum. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27(3):623–633, 2000.

Jamie Woych, Alonso Ortega Gurrola, Astrid Deryckere, Eliza CB Jaeger, Elias Gumnit, Gianluca Merello, Jiacheng Gu, Alberto Joven Araus, Nicholas D Leigh, Maximina Yun, et al. Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. *Science*, 377(6610):eabp9186, 2022.

Yanbo Xue, Le Yang, and Simon Haykin. Decoupled echo state networks with lateral inhibition. *Neural Networks*, 20(3), 2007.

Guangyu Robert Yang and Manuel Molano-Mazón. Towards the next generation of recurrent network models for cognitive neuroscience. *Current opinion in neurobiology*, 70:182–192, 2021.

Hang Zhang, Xiangjuan Ren, and Laurence T Maloney. The bounded rationality of probability distortion. *Proceedings of the National Academy of Sciences*, 117(36): 22024–22034, 2020.

Ruixun Zhang, Thomas J Brennan, and Andrew W Lo. The origin of risk aversion. *Proceedings of the National Academy of Sciences*, 111(50):17777–17782, 2014.

Zhewei Zhang, Zhenbo Cheng, Zhongqiao Lin, Chechang Nie, and Tianming Yang. A neural network model for the orbitofrontal cortex and task space acquisition during reinforcement learning. *PLOS Computational Biology*, 14(1):e1005925, 2018.

Changsong Zhou, Lucia Zemanová, Gorka Zamora, Claus C Hilgetag, and Jürgen Kurths. Hierarchical organization unveiled by functional connectivity in complex brain networks. *Physical review letters*, 97(23):238103, 2006.

Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560, 1997.