



**HAL**  
open science

# From State Transitions to Sensory Regularity: Structuring Uninterpreted Sensory Signals from Naive Sensorimotor Experiences

Loïc Goasguen, Jean-Merwan Godon, Sylvain Argentieri

► **To cite this version:**

Loïc Goasguen, Jean-Merwan Godon, Sylvain Argentieri. From State Transitions to Sensory Regularity: Structuring Uninterpreted Sensory Signals from Naive Sensorimotor Experiences. 2022. hal-03537409v3

**HAL Id: hal-03537409**

**<https://hal.science/hal-03537409v3>**

Preprint submitted on 28 Sep 2022 (v3), last revised 10 Mar 2023 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# From State Transitions to Sensory Regularity: Structuring Uninterpreted Sensory Signals from Naive Sensorimotor Experiences

Loïc Goasguen\*, Jean-Merwan Godon\* and Sylvain Argentieri

**Abstract**—How could a naive agent build some internal, subjective, notions of continuity in its sensorimotor experiences? This is a key question for all sensorimotor approaches to perception when trying to make them face realistic interactions with an environment, including noise in the perceived sensations, errors in the generation of motor trajectories, or uncertainties in the agent’s internal representation of this interaction. This paper proposes a detailed formalization, but also some experimental assessments, of the structure a naive agent can leverage from its own uninterpreted sensorimotor flow to capture a subjective sensory continuity, making it able to discover some notions of closeness or regularities in its experience. The precise role of the agent’s actions is also questioned w.r.t. the spatial and temporal dynamics of its exploration of the environment. On this basis, the previous authors’ contribution on sensory prediction is extended to successfully handle noisy data in the agent’s sensorimotor flow.

**Index Terms**—Sensorimotor contingencies theory, topological grounding, sensory regularities, uninterpreted sensory signals.

## I. INTRODUCTION

It is certainly the case that we deem our sensory experience to be “continuous”. Indeed, one crucial property of many psychological perceptual processes is that they generally *seem* continuous [1]; in point of fact, this intuition is strong enough that it is the converse situations where it visibly is not that earn explicit mentions, such as that of Categorical Perception [2], [3]. However, such continuity does not trivially follow from our knowledge of how perceptual processes are materially –e.g. neurally– mediated [4], [5], [6]. In the instance of visual perception, for example, it is known that the eye only acquires very partial snapshots of visual information due to the sparse layout of discrete photoreceptors on its retina as well as the typical trajectories of ocular saccades.

Nevertheless, the continuity of perception subjectively experienced by sensorimotor agents is undeniably useful, allowing for the formulation and exploitation of several powerful ideas. One such idea, for instance, is that of inter and extrapolation. If an agent hopes to infer properties of an unknown situation from a structure it has learned from previous experiences, this agent should have a way to quantify in what way this new experience relates to the data it already knows. One very common way to deal with this is thus to *a priori* assign close properties (e.g. evaluated in terms of distances between sensory signals features, proximity between spatial positions) to experiences that are themselves alike (e.g. by characterizing

similar physical properties in the environment, or by rating the system ability to achieve its task): the agent should then have the capabilities to distinguish “similar” things, be it external objects (e.g. a cymbal emitting a sound), sensory attributes (e.g. the intensity, or the tone of the same cymbal), or even sensorimotor capabilities (e.g. the association between these attributes and the action actually performed by the agent to emit the sound from the cymbal). These capacities may in turn provide grounds for the emergence of its felt continuity of perception: in the end, the agent should then be able to assert that “Red is closer to Pink than it is to Blue, and it is certainly closer to Blue than it is to the sound of a bell” [7]. Such closeness properties are usually leveraged in robotic settings through the well-known mathematical notion of continuity of maps  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  since the data available to the robotic agent is usually represented numerically. More generally, the modern examination of continuity and related problems is the subject of *topology* [8], a field of mathematics which is precisely devoted to the study of what it means for something to be continuous. This field has indeed proved a powerful tool for bootstrapping [9], or for modeling geometric ideas in several sensorimotor works [10], [11], [12], in particular those that attempt to internally establish properties of external space [13]. Such approaches allow, e.g., motion planning in the internal sensorimotor body representation of an agent through the generation, by interpolation, of continuous motor trajectories [14], or the emergence of a topological representation of the sensor poses from the sensorimotor flow [15]. But importantly, while most of these works are rooted in generic topological intuitions, they all end up exploiting a discrete setup for which most of the topological structures are useless. Indeed, most modern robotic setups rely on discrete time computations for which we can define other tools like distances based on similarities or correlations between elements in the agent’s sensorimotor flow. Then, should we want a naive agent to make some kind of judgement about discrete samples by way of its *subjective* sense of continuity, then this sense cannot be entirely grounded in topology; in particular, it cannot be reduced to that of formal continuity. As a consequence, the (almost) only assessments one might provide a naive agent with are entirely categorical: it should then only be able to perform comparisons at a “symbolic” level, denoted by a strict equality operator between e.g. sensory values. While the previous cited contributions certainly prove that these operations allow for the extraction of interesting features or meaningful internal representations from a naive form of sensorimotor flow, they also share limitations related to the absence of the

\* Loïc Goasguen and Jean-Merwan Godon have both equally contributed to this paper. All authors are with Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, ISIR, F-75005 Paris, France.

93 aforementioned “closeness” concept: what about their robust-  
 94 ness w.r.t. noise, imperfect repetition of motor paths, etc.? The  
 95 very same limitation is also shared by the previous work by  
 96 the authors [16]; in this contribution, the interlink between  
 97 motor actions and sensory prediction is explored, through  
 98 the demonstration of the existence of a group isomorphism  
 99 between them. But predicting the sensory outcome of an action  
 100 is only accessible to the agent by detecting the exact shift  
 101 of values inside its own sensor array. Endowing the agent  
 102 with some internal notion of sensory closeness would then  
 103 make it able to assess its own prediction, and more generally,  
 104 might allow these sensorimotor approaches to perception –so  
 105 far mainly restricted to simulated territories– to deal with more  
 106 realistic conditions.

107 Importantly, most of the previously cited contributions also  
 108 claim to deal with uninterpreted sensory signals. But in these  
 109 works, the form assumed by the signal (and the expected  
 110 transformations thereof) is usually known and leveraged by  
 111 the agent; what it ignores instead is *how* these signals relate  
 112 to the sensorimotor interaction. Then, using a priori distances,  
 113 metrics, and similarities, maps or representations of the agent’s  
 114 sensorimotor interaction with a generally unknown environ-  
 115 ment are built [17]. In this paper, however, no “natural” metrics  
 116 nor algebraic operations on the symbols perceived by the  
 117 agent are used –in a similar way to [18] where a less formal  
 118 approach is proposed–, contrary to our understanding of the  
 119 usual numeric values. This is generally made manifest with the  
 120 choice to assume that states are coded as numeric values (or  
 121 tuples thereof), and of special influence with that of whether  
 122 to use natural (possibly topological, as previously outlined)  
 123 structures of  $\mathbb{R}^N$ . Thus, most developments which try to  
 124 achieve robustness and scalability do so via extrapolation and  
 125 clustering [12], [14], while [19] goes a bit further by evaluating  
 126 sensory states similarities by their transition probabilities, but  
 127 for object identification in a sensorimotor context. Never-  
 128 theless, as already argued, most of these techniques require  
 129 referring to preexisting external metrics, which constitutes  
 130 assumption about a priori knowledge we would like to avoid.

131 In this paper, we then propose to examine how some notion  
 132 of closeness –that we could also relate to some internal  
 133 notion of *subjective* continuity– in sensorimotor experiences  
 134 can emerge from uninterpreted sensory information for a naive  
 135 agent operating in discrete time. To that end, some formal  
 136 considerations are first introduced in §II. After evaluating a  
 137 purely topological approach, a metric approach is proposed  
 138 instead and the probability of transition between sensory  
 139 symbols is used to define some appropriate notion of sensory  
 140 distance. On this basis some simple simulations are introduced  
 141 in Section III to illustrate how an agent could leverage some  
 142 structure by simply judging if its sensory observations are  
 143 close or not. This is illustrated for visual perception through  
 144 the building by a naive agent of the grayscale or some RGB  
 145 color model. Then, the role of the agent’s action in this  
 146 framework is questioned in §IV. More precisely, the spatial  
 147 and temporal dynamics of the agent’s exploration is shown  
 148 critical to obtain a meaningful and useful structure of its own  
 149 sensory symbols. Next, some experiments initially proposed  
 150 in [16] are reproduced in Section V to illustrate how the  
 151 proposed framework could allow an agent to actually build

some sensory prediction functions even in the presence of  
 sensory noise. Finally, a conclusion ends the paper.

## II. TOWARDS A TOPOLOGY OF SENSORY VALUES

This first section aims at defining a topology of sensory val-  
 ues, built on the basis of the agent’s sensorimotor experience.  
 After a short subsection devoted to the required definitions  
 and notations, a time variable is added to the formalism in  
 the second subsection, so as to account for the explicit time  
 dependency of the agent’s experience, allowing us to introduce  
 a first time-inherited topology. While being possibly sufficient,  
 arguments for the introduction of an explicit metric are then  
 discussed. The third subsection thus proposes the definition  
 of an internal probabilistic metric and highlights the benefits  
 and limits of the proposed approach. Section III then exploits  
 these elements in a simple experimental framework to illustrate  
 these elements and demonstrate their actual exploitation.

### A. A short reminder on notations

Let us consider in the following an agent endowed with  
 motor and sensory capabilities. Its internal sensorimotor con-  
 figuration is classically noted as  $(\mathbf{m}, \mathbf{s})$ , where  $\mathbf{m} \in \mathcal{M}$  (resp.  
 $\mathbf{s} \in \mathcal{S}$ ) represents the agent’s internal motor (resp. sensory)  
 configuration as an element of its corresponding motor  $\mathcal{M}$   
 (resp. sensory  $\mathcal{S}$ ) set. As shown in [16], the agent’s motor  
 description can be enriched from  $\mathbf{m} \in \mathcal{M}$  to  $\mathbf{b} \in \mathcal{B}$ , where  
 $\mathbf{b} = (\mathbf{m}, \boldsymbol{\tau})$  depicts the absolute agent’s motor configuration.  
 $\mathbf{b}$  is made of the agent’s internal (and thus known to it)  
 motor configuration  $\mathbf{m}$  and of its absolute external (and thus  
 unknown to it) pose  $\boldsymbol{\tau}$  in its ambient space. Importantly,  
 as discussed in [16], switching the motor description from  
 $\mathbf{m}$  to  $\mathbf{b}$  allows us to keep a functional relationship between  
 motor and sensory data, even in the case where the agent  
 can freely move in its environment. But while the agent has  
 no direct access to  $\mathbf{b}$ , it can apply some *motor actions*  $a$  on  
 $\mathbf{b} = (\mathbf{m}, \boldsymbol{\tau})$  to go to configurations  $\mathbf{b}' = (\mathbf{m}', \boldsymbol{\tau}') = a\mathbf{b}$ : the  
 agent knows instead how to *move* in  $\mathcal{B}$ . This capability will  
 be exploited later to apply the following developments to get  
 an internal assessment of sensory regularity, see §V. Next, the  
 environment state is characterized as a function  $\epsilon : \mathcal{X} \rightarrow \mathcal{P}$ ,  
 i.e. as a state  $\epsilon \in \mathcal{E}$  linking the ambient geometrical space  $\mathcal{X}$   
 in which sensorimotor experiences occur (classically endowed  
 with some rigid transformations group  $\mathcal{G}(\mathcal{X})$ ) to the set of  
 the physical properties  $\mathcal{P}$  observable by the agent, where  $\mathcal{E}$   
 denotes the set of environmental states. Then,  $\epsilon(\mathbf{x})$  represents  
 the observable physical properties at point  $\mathbf{x} \in \mathcal{X}$ . On the  
 basis of the previous definitions, we can now define the  
 sensorimotor map  $\psi$  as the function  $\psi : \mathcal{B} \times \mathcal{E} \rightarrow \mathcal{S}$ , such  
 that  $\mathbf{s} = \psi(\mathbf{b}, \epsilon)$ . We can notice here that the sensorimotor  
 law does not explicitly depend on time, as is the case of most  
 other contributions in the fields [10], [20], [14]. We will now  
 enrich this formalization with an explicit time dependency.  
 It will then constitute our gateway towards continuity in the  
 sensory experience of the agent. Much like in J. Elman’s  
 famous 1990 paper [21] where words are not considered as  
 preexisting categories but more as emergent features in the  
 latent structure of sentences along time [22]. Similarly, the  
 topology of sensory symbols can be considered as the latent  
 structure between them through the sensorimotor experience  
 along time.

210 *B. All is well in continuous land*

211 1) *Introducing time in the sensorimotor experience:* The  
212 definitions we recalled in the previous subsection actually  
213 described *snapshots* of the agent’s sensorimotor interaction.  
214 Nevertheless, these can be easily enriched with an explicit  
215 dependency of the various states with a time variable  $t \in \mathcal{T}$ .  
216 Thus, the environmental state  $\epsilon \in \mathcal{E}$  can now be written

$$\begin{aligned} \epsilon: \mathcal{T} \times \mathcal{X} &\rightarrow \mathcal{P} \\ (t, \mathbf{x}) &\mapsto \epsilon(t, \mathbf{x}). \end{aligned} \quad (1)$$

217 With this notation, we can express an instantaneous snapshot  
218 of the environmental state as the partial function

$$\epsilon_t: \mathbf{x} \in \mathcal{X} \mapsto \epsilon_t(\mathbf{x}) = \epsilon(t, \mathbf{x}) \in \mathcal{P}. \quad (2)$$

219 Therefore, any temporal succession of environment states can  
220 be described as a trajectory

$$\gamma_\epsilon: t \in \mathcal{T} \mapsto \epsilon_t \in \mathcal{E}. \quad (3)$$

221 Correspondingly, the agent’s *absolute* configuration trajec-  
222 tories and sensory ones are respectively denoted by

$$\gamma_{\mathbf{b}}: t \in \mathcal{T} \mapsto \mathbf{b}_t \in \mathcal{B}, \quad (4)$$

223 and

$$\gamma_{\mathbf{s}} = \gamma_{\mathbf{b}, \epsilon}: t \in \mathcal{T} \mapsto \mathbf{s}_t = \psi(\gamma_{\mathbf{b}}(t), \gamma_\epsilon(t)) \in \mathcal{S}. \quad (5)$$

224 In the following, we will consider a particular subset of  
225 such temporal environmental, motor, and sensory trajectories  
226 representing the set of *effectively valid* trajectories. We thus  
227 instead restrict  $\epsilon_t \in \mathcal{E}_{\mathcal{T}} \subset \mathcal{F}(\mathcal{T}, \mathcal{E})$  so as to include possible  
228 external constraints on the succession in time of physical  
229 properties in the agent’s environment. In the same vein, one  
230 defines  $\gamma_{\mathbf{b}} \in \mathcal{B}_{\mathcal{T}}$ , where  $\mathcal{B}_{\mathcal{T}}$  is the set of all effectively  
231 performable motor configurations, possibly allowing to capture  
232 e.g. limitations on velocity and their smoothness as actuated by  
233 the agent. Consequently, the effectively valid sensory trajec-  
234 tories  $\gamma_{\mathbf{s}}$  lie in  $\mathcal{S}_{\mathcal{T}}$ , with a natural mapping  $\mathcal{S}_{\mathcal{T}} \hookrightarrow \mathcal{B}_{\mathcal{T}} \times \mathcal{E}_{\mathcal{T}}$ .

235 2) *Towards a sensory topology:* Let us now get back to  
236 the intuition of the sensory experience being continuous, as  
237 discussed in the introduction of this paper. More precisely, this  
238 continuity is that of the agent’s sensory experience unfolding  
239 with the time  $\mathcal{T}$  during which it occurs. In (purely) topo-  
240 logical settings, an argument examined e.g. in [23] shows that  
241 searching for (*formal*) continuity of the  $\gamma_{\mathbf{s}}$  sensory experiences  
242 is entirely dual to searching for topological constraints on  
243 the sensory values  $\mathbf{s} \in \mathcal{S}$ . These two viewpoints intersect  
244 at the *final topology* of the  $\gamma_{\mathbf{s}}$  [8], a topology on  $\mathcal{S}$  which  
245 precisely encodes which structural constraints on the  $\mathbf{s}$  sensory  
246 values is needed to make (all) the  $\gamma_{\mathbf{s}}$  experiences continuous.  
247 While this final topology seems to solve –at least from a  
248 purely topological point of view– the initial problem, we have  
249 to keep in mind that most robotic setups rely on discrete  
250 time computations. The resulting final topology thus makes  
251  $\mathcal{S}$  discrete. Intuitively, this occurs because if the agent only  
252 experiences jumps in times such that no instant follows  
253 continuously from the previous one, then it does not need  
254 to introduce new continuities in its sensations to make their  
255 succession continuous. So how can we solve this issue? In the  
256 following subsection, we propose to switch to metric geometry,  
257 which, although less general, might be better suited.

*C. Introduction of a statistical sensory metric*

258

Introducing corresponding metric considerations, however, 259  
raises new issues: given an abstract sequence of points in 260  
a (metrized) point cloud, how can we determine whether it 261  
represents a regular/continuous trajectory? For example, how 262  
can we decide that a jump in values across a distance of e.g. 5 263  
units corresponds to a *regular* transition, or instead represents 264  
a break in continuity? Without *a priori* assumptions about the 265  
expected reasonable dynamics of the experience, it seems these 266  
numbers are entirely arbitrary, and related to some *external* 267  
knowledge that we want the agent to do without. Instead, 268  
we propose to define a statistical sensory metric, for which 269  
the agent ought to set to zero any distance between sensory 270  
values that *immediately* (and not *continuously*) follow one 271  
another. Thus, the temporal length between successive sensory 272  
samples is now central to how the agent perceives them. 273  
Consequently, we should first assume that the agent is able 274  
to compute distances (or durations) between two timesteps in 275  
 $\mathcal{T}$ . On this basis, we will assume in all the following that the 276  
laws of the sensorimotor experiences the agent can observe 277  
are *time homogeneous*. This hypothesis then indicates that no 278  
statistical measurement the agent can empirically obtain from 279  
its sensorimotor experience may depend on the absolute value 280  
of the timestep indexing its interaction. In particular, it should 281  
be a natural consequence of the particular choice of timestep 282  
being an entirely external convention, implementing a sort of 283  
independence of choice of reference. 284

Let us now define the likelihood  $P_{s'|s}$  over all experiences 285  
that the sensory value  $s'$  immediately follows  $s$  in the senso- 286  
rimotor flow of the agent along 287

$$P_{s'|s} = \mathbb{P}(\gamma_{\mathbf{s}}(t+1) = s' \mid \gamma_{\mathbf{s}}(t) = s). \quad (6)$$

Importantly, from the previous time homogeneity assumption, 288  
 $P_{s'|s}$  does not depend on the current time  $t$  it is computed. 289  
From there and following the intuition that “closeness” of sen- 290  
sory values  $\mathbf{s}$  and  $\mathbf{s}'$  should increase whenever the probability 291  
of the transition  $\mathbf{s} \rightarrow \mathbf{s}'$  does, we propose to define a simple 292  
metric prototype via 293

$$\delta_f(\mathbf{s}, \mathbf{s}') = f(P_{s'|s}) \quad \forall \mathbf{s}, \mathbf{s}' \in \mathcal{S}, \quad (7)$$

where  $f$  should verify the two conditions: 294

- 1)  $f: [0; 1] \rightarrow \mathbb{R}_+$ :  $f$  only needs to map probabilities in 295  
[0; 1] to nonnegative values, i.e. dissimilarity values; 296
- 2)  $f$  is non-increasing: probable transitions (i.e.  $P_{s'|s}$  close 297  
to 1) should result in low dissimilarities. 298

These conditions do not make  $\delta_f$  a metric since it only verifies 299  
the non-negativity property. We therefore extend it via minimal 300  
path considerations, i.e., by defining a distance  $d_f$ . Let  $\mathcal{R}^{\mathbf{s}, \mathbf{s}'}$  301  
be the set of all paths from  $\mathbf{s}$  to  $\mathbf{s}'$ , with 302

$$\langle \mathbf{s} = \mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \dots, \mathbf{s}^{(k-1)}, \mathbf{s}^{(k)} = \mathbf{s}' \rangle \in \mathcal{R}^{\mathbf{s}, \mathbf{s}'}. \quad (8)$$

We can then define  $d_f$  along 303

$$d_f(\mathbf{s}, \mathbf{s}') = \inf \mathcal{R}^{\mathbf{s}, \mathbf{s}'}. \quad (9)$$

This in turn enforces the properties of *triangular inequality* 304  
and *reflexivity*. In the case where  $\mathcal{S}$  is finite, this reduces 305  
to the familiar computational form of finding minimal paths 306  
on a finite graph with nonnegative weights (corresponding 307

to the  $\delta_f(\mathbf{s}, \mathbf{s}')$  edge from  $\mathbf{s}$  to  $\mathbf{s}'$ ). It should also be noted that this does *not* guarantee *symmetry* at its core because  $P_{\mathbf{s}'|\mathbf{s}}$  may differ from  $P_{\mathbf{s}|\mathbf{s}'}$ . Then the  $\delta_f$  weights naturally define a *directed* graph (*digraph*), which does not impair the search for minimal paths but does however, lead to a non-symmetric  $d_f$  function. While there exist several ways to obtain a closely related undirected graph from any given digraph, we hypothesize instead that symmetry should occur as a contingency of the sensorimotor exploration in most real world examples. Therefore, we do not enforce such corrections for now and will instead assess this hypothesis in the resulting graph.

### III. BUILDING THE SENSORY TOPOLOGY FROM STATISTICS

The previous section was devoted to the mathematical roots of the approach. We will now illustrate how these points can be exploited inside a simple experimental framework which could allow a naive agent to leverage a structure on its sensory signals from its own observations. To begin with, a detailed description of the simulation setup is proposed. On this basis, two main experiments are conducted: the first one deals with the construction of a probabilistic sensory metric and the corresponding low-embedding representation for a grayscale camera sensor; the second one extends the reasoning to a more complex representation when using RGB image sensors.

#### A. Experimental setup and sensory distance estimation

1) *Experimental setup*: In all the following, we consider an agent endowed with a camera sensor observing a 3D scene. Since we are for now dealing with sensory values and their transitions only, the visual perception is basically simulated by playing a video file  $\mathbf{v}[n]$  of size  $W \times H$ , where  $n$  represents the video frame number. This is a (temporary) very restrictive setup, which will be enriched later when discussing the influence of the movement of the agent (see §IV). Also, the experience occurs in discrete time, for which each timestep verifies  $t = t_n = nT_s$  with  $T_s$  the sampling period. In practice, we have  $\mathbf{v}[n] = (v_{ij}[n])_{i,j}$ , with  $i \in [0; W-1]$ ,  $j \in [0; H-1]$ , and where  $v_{ij}[n]$  depicts the pixel value of the video at frame  $n$ , row  $i$  and column  $j$ . Each pixel  $v_{ij} = (R_{ij}, G_{ij}, B_{ij})$  is represented as a traditional color tuple  $\in [0; 255]^3$ . The agent's sensory state  $\mathbf{s}[n]$  is then simulated by applying some instantaneous function  $g : [0; 255]^3 \rightarrow \mathcal{S}$  to the video, i.e.

$$\mathbf{s}[n] = (s_{ij}[n])_{i,j}, \text{ such that } s_{ij}[n] = g(v_{ij}[n]), \quad (10)$$

where  $s_{ij}[n]$  represents the  $(i, j)$  sensel value at time  $n$ , row  $i$  and column  $j$  of the agent's camera sensor. Introducing  $g(\cdot)$  in (10) allows to explain formally how a physical state of the environment (which can be envisaged here as the pixel values of the video) is turned into the internal sensory state of the agent. But one has to keep in mind that the agent does not know the relation (10), it does not even have any knowledge about the meaning of these numerical values: they are only *uninterpreted symbols* to it, with no a priori structure, order, nor any way to actually *compare* them. In addition, the set  $\mathcal{S}$  may well be isomorphic to the set of actual pixel values, but there may also have a lower number  $S$  of symbols than pixel values, resulting in a compressed representation. Without loss

of generality,  $\mathcal{S}$  will then be defined as the finite set of positive integers  $\{0, \dots, S-1\}$  with  $S = \text{Card}(\mathcal{S})$ , where each sensory symbol  $\mathbf{s}_k \in \mathcal{S}$  can equally be written directly as the integer  $k$ , and we will adopt a traditional  $s_{ij} \in [0; S-1]$  coding convention for the numerical values of each  $(i, j)$  sensel, with  $S = 256$  for traditional camera sensors. As outlined in §II-C, it is then proposed to look at the relationship between those  $S$  uninterpreted (numerical) symbols through the statistics of their transitions. Let us now detail how these transitions are actually captured.

2) *Description of the experiment*: In all the following, we will assume that all  $W \times H$  agent's sensels contribute equally to the building of the same representation, i.e., all sensels share the same excitation function linking the environment state to the agent's sensations as written in Equation (10). Then, we define a  $S \times S$  matrix  $M = (m_{kl})_{k,l}$  counting all the transitions of sensel values along observations, with

$$m_{kl}[n+1] = m_{kl}[n] + \sum_{i,j} \zeta_{kl}(i, j)[n], \quad (11)$$

with  $(i, j) \in [1; W \times H]^2$ ,  $m_{kl}[0] = 0$ , and  $k, l$  both represent two symbols in  $\mathcal{S}$  (that is, sensor output values  $\mathbf{s}_k$  and  $\mathbf{s}_l \in \mathcal{S}$ ).  $\zeta_{kl}(i, j)[n]$  aims to capture the existence of a change of value of the  $(i, j)$  sensel from value  $k$  at time  $n$  to value  $l$  at time  $n+1$ , i.e.

$$\zeta_{kl}(i, j)[n] = \begin{cases} 1 & \text{iff } s_{ij}[n] = k \text{ and } s_{ij}[n+1] = l, \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

From (11), we can then compute the probability of transition of sensels values gathered in a  $S \times S$  matrix  $P = (p_{kl})_{k,l}$  with

$$p_{kl}[n] = \frac{m_{kl}[n]}{\sum_{q=0}^{S-1} m_{kq}[n]} \quad (13)$$

the probability at time  $n$  for any sensel to see its value changing from symbol  $k$  to  $l$ . Obviously,  $p_{kl}[n]$  is expected to converge towards  $P_{\mathbf{s}_l|\mathbf{s}_k}$  as time  $n$  tends to infinity. Then, once the estimation of the matrix  $P$  has converged after a fixed number frames  $N$ , it is turned into a  $S \times S$  metric prototype matrix  $\Delta = (\delta_{kl})_{k,l}$  according to Eq. (7) where  $f = -\log^1$  is selected, with

$$\delta_{kl} = -\log(p_{kl}[N]). \quad (14)$$

Again, any function verifying the two conditions in §II-C could have been selected. Then, Dijkstra's algorithm [24] is applied to the  $\Delta$  matrix along Eq. (9) to produce the  $S \times S$  distance matrix  $D = (d_{kl})_{k,l}$ , providing the agent with the result metric  $d$  we set out to discover

$$d_f(\mathbf{s}, \mathbf{s}') = d_{-\log}(\mathbf{s}_k, \mathbf{s}_l) = d_{kl}, \quad (15)$$

which is finally visualized in 2D or 3D through a multi-dimensional scaling projection method (MDS [25], [26], [27] or ISOMAP [28]).

#### B. Results for a grayscale perception

The  $W \times H = 856 \times 480$  video used to conduct the experiments comes from a slightly stabilized camera filming

<sup>1</sup>If a probability of transition is equal to 0, the corresponding distance is set to NaN by convention.

404 an evening walk in Midtown New York City in the rain<sup>2</sup>. It  
 405 consists of a natural city scene filmed in real time from a  
 406 first-person point of view. A grayscale (cropped) preview of  
 407 the video is shown in Figure 1a. It is clear that this environ-  
 408 ment exhibits some nice local temporal and spatial continuity  
 409 properties: the values of each pixel change smoothly in time,  
 410 while local pixel values are highly correlated. While these are  
 411 some nice properties to illustrate the building of the sensory  
 412 topology from statistics, the importance and formalization of  
 413 these hypotheses w.r.t. the agent’s movement capabilities is  
 414 discussed in §IV.

415 To begin with, we will consider a function  $g$ , mapping the  
 416  $(R_{ij}, G_{ij}, B_{ij})$  color coding of the video pixels  $v_{ij}$  to the  
 417 sensel values  $s_{ij} \in \llbracket 0; 255 \rrbracket$  of the agent, such that

$$s_{ij} = g(v_{ij}) = h(\text{round}(\text{mean}(R_{ij}, G_{ij}, B_{ij}))), \quad (16)$$

418 where  $h$  is a function that can be tuned to artificially modify  
 419 the agent’s perception. Note that  $g$  acts here like an exci-  
 420 tation function, and is thus supposed to be identical for all  
 421 sensels. Two cases for  $h$  are discussed in the following: either  
 422  $h() = \text{id}()$  in §III-B1, corresponding to the case where the  
 423 agent’s grayscale perception exactly matches the grayscale  
 424 version of the video, or  $h() = \text{sawtooth}()$  for which the  
 425 perception is altered on purpose to exhibit the folding of the  
 426 agent’s internal representation between black and white pixel  
 427 values in the video, as detailed in §III-B2.

428 1) *First case:  $h()$  is the identity function:*

429 a) *Estimation of the probability of transition between*  
 430 *symbols:* Since  $h() = \text{id}()$  in Eq. (16), the agent’s sensory  
 431 values are made of  $S = 256$  uninterpreted symbols, whose  
 432 values along frames can be used to compute their probability  
 433 of transition along Equation (13). The resulting  $S \times S$  matrix  
 434  $P$  is shown in Figure 1b and 1e after  $n = 5$  and  $n = 10^4$   
 435 successive sensory transitions respectively. Note that the  $S$   
 436 symbols are ordered in the figure according to their numerical  
 437 values: this is something the agent cannot actually do for now,  
 438 but this ordering has no effect on the reasoning and helps in  
 439 understanding the process. From Figures 1b and 1e, we can  
 440 see that the most probable transitions are all placed along the  
 441 diagonal of the matrix  $P$ , meaning that the most probable  
 442 sensory output at the next time step is the very same symbol,  
 443 even at the very beginning of the experiment with  $n = 5$ .  
 444 Further, the *a priori* ordering of symbols allows to observe that  
 445 the diagonal is thick and fades away as the symbols values are  
 446 distant: this clearly indicates that the most probable transitions  
 447 are the ones to symbols that have *close colors, from an external*  
 448 *point of view* (again, the *a priori* ordering is unknown to  
 449 the agent). Conversely, the least probable transitions are the  
 450 ones to *distant* symbols. Those results are in accordance with  
 451 the intuition that close time intervals lead to close sensory  
 452 outputs, and that some regularity of the sensory experience  
 453 has been captured. Note that since the probability estimation  
 454 is evaluated on occurrences, the case where no transitions at  
 455 all between two symbols are observed leads to a probability  
 456 of 0 (represented in white in Figure 1b); this appears at the  
 457 beginning of the experiment only (see Figure 1e for  $n = 10^4$ )  
 458 and mainly concerns *distant* symbols with a very low transition  
 459 probability, i.e., in the two corners of Figure 1b.

<sup>2</sup>[https://youtu.be/eZe4Q\\_58UTU](https://youtu.be/eZe4Q_58UTU) by courtesy of Nomadic Ambience.

b) *Computation of the distance matrix:* On the basis  
 of the previous probability of transitions between symbols,  
 we can compute the metric prototype in the form of the  
 $S \times S$  matrix  $\Delta$  whose elements are given by Eq. (14).  
 Then, Dijkstra’s algorithm [24] is performed on  $\Delta$  to obtain  
 the  $S \times S$  distance matrix  $D$ . The resulting matrix  $D$  is  
 represented in Figure 1c and 1f for  $n = 5$  and  $n = 10^4$   
 respectively. Obviously, one should note that when direct  
 transitions between symbols are missing in  $P$  (and thus in  $\Delta$ )  
 as shown in Figure 1b, Dijkstra’s algorithm will nonetheless  
 generally find an alternate path towards those symbols by  
 finding adequate successive transitions; consequently the  $D$   
 matrix is expected to be fully defined (i.e. with all coefficients  
 finite) as long as the agent has experienced enough sensory  
 symbols transitions. This is exactly what is shown in Figure 1c,  
 where the corresponding distance matrix  $D$  shows distances  
 between all sensory symbols, while transitions between some  
 of them have not been directly observed yet. We can also  
 see from both Figures 1c and 1f that previous low transition  
 probabilities are now associated with high distances (and vice  
 versa). In addition, we recognize the same diagonal pattern,  
 which now corresponds to low distances. We can also see  
 that  $D$  is *almost* symmetric, except in the corners, where lie  
 most of the high distances, corresponding to the least probable  
 transitions of sensory symbols. This is not an encoded property  
 of the agent’s experience but instead seems to appear as  
 a contingency of the sensorimotor exploration, as outlined  
 in §II-C. Finally, a qualitative comparison between the two  
 $D$  matrices obtained at the beginning (Figure 1c) and at  
 the end (Figure 1f) of the experiment shows that the very  
 same structures (symmetry, diagonal pattern) are captured very  
 quickly. This is certainly thanks to the identical contribution  
 of all pixels to the building of the same statistic, as one time  
 step actually captures  $W \times H \approx 4.10^5$  sensory transitions.

c) *Visualization of the representation:* Finally, we can  
 qualitatively assess the shape of the captured sensory symbols  
 topology by projecting the resulting distance matrix  $D$  into a  
 space of lower dimension. The 2D visualization of the matrix  
 $D$  through a MultiDimensional Scaling (MDS) projection is  
 represented in Figures 1d and 1g. Note that such a method  
 requires the input matrix to be symmetric; hopefully, we  
 qualitatively showed it was almost the case so that MDS can be  
 actually applied on the symmetrized matrix  $1/2 \times (D + D^T)$ . In  
 both Figures 1d and 1g, each circle represents a single symbol  
 where the inner color corresponds to the color perceived from  
 an external point of view (color that also matches the classical  
 gray-level scale in this case, since  $f = \text{id}()$ ). We can see from  
 this representation that the obtained manifold is almost one-  
 dimensional and captures the classical gray scale from white to  
 black in a continuous manner, even at the very beginning of the  
 experiment. This can be evaluated by looking for the 2 nearest  
 neighbors of each symbol in the internal metric (i.e. with the  
 neighbors computed on  $D$  and not on the representation); these  
 neighbors are then linked together in the projection by a line  
 drawn in the figure. Browsing the manifold by following these  
 lines allows to go from white (coded as the number 255)  
 to black (coded as a 0) almost without any discontinuity in  
 the symbol order at the end of the experiment. Interestingly,  
 we can see that the projection obtained at the early stage of

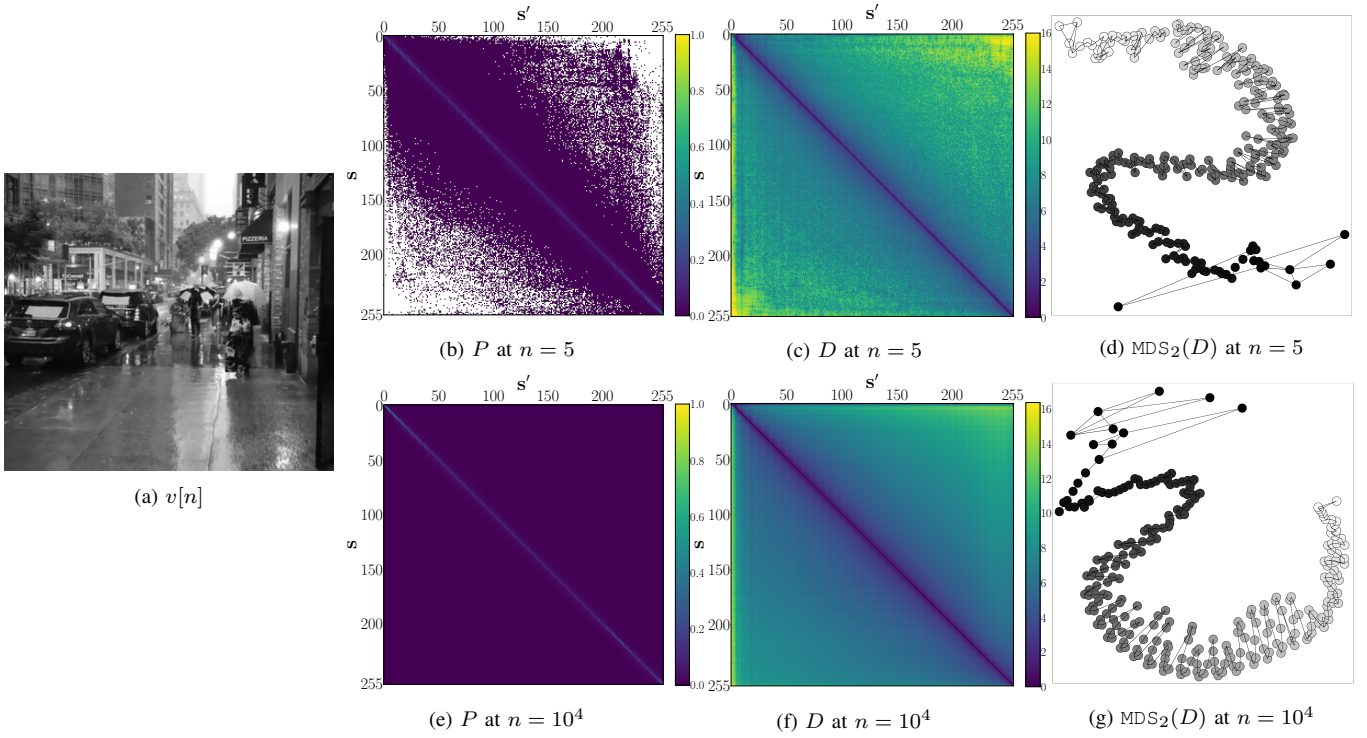


Figure 1: Building of the internal organization of sensory values. (a) Grayscale version of one frame of the video used in the experiment. (b)(e) Estimated probability matrix  $P$  at  $n = 5$  and  $n = 10^4$ , i.e., at the very beginning of the experiment. (c)(f) Estimated distance matrix  $D$  at  $n = 5$  and  $n = 10^4$ , i.e., at the end of the experiment. (d)(g) Corresponding low-dimensional embedding of  $D$  at  $n = 5$  and  $n = 10^4$ : we can see the intuitive grayscale organization of pixel values, discovered by the agent from its sensory values transitions.

519 the experiment already exhibits a one dimensional manifold,  
 520 with a thicker and less organized ordering of symbols. Again,  
 521 the contribution of all sensels to the same statistic certainly  
 522 explains this nice quick convergence of the representation.  
 523 Thus, from the final graph, we can conclude that the agent  
 524 has been able, starting only from the probability of transition  
 525 between uninterpreted sensory symbols, to discover the gray  
 526 level scale. Such a capability will be further exploited for  
 527 different applications, like sensory prediction, see §V.  
 528 2) 2nd case:  $h()$  is a sawtooth function: We will now  
 529 consider a case where the agent's sensory output does not  
 530 exactly match the original grayscale world as per Eq. (16),  
 531 where  $h() = \text{sawtooth}()$  is defined along

$$\text{sawtooth}(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 127 \\ 2(x - 128) & \text{otherwise,} \end{cases} \quad (17)$$

532 for  $x \in \llbracket 0; 255 \rrbracket$  only. With such a change, a single internal  
 533 sensory symbol (e.g., 54) will now correspond to two possible  
 534 world grayscale values (27 and 155). Intuitively, such a change  
 535 is expected to *create* continuity that does not exist initially  
 536 between symbols through a closer proximity between values  
 537 representing dark and light shades. The previous process  
 538 is then repeated and the resulting 2D MDS embedding is  
 539 depicted in Figure 2: as expected, we identify a looping  
 540 monodimensional manifold. In the figure, each sensory symbol  
 541 is depicted as a circle whose color represents its *internal*  
 542 coding (i.e., a numerical value from 0 to 254 with a step of 2),  
 543 represented as grayscale values for convenience. This color no  
 544 longer matches the grayscale values of the world it represents  
 545 because of the introduction of the sawtooth function. But the

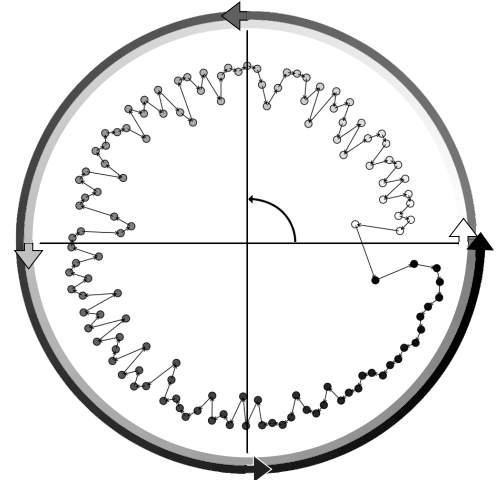


Figure 2: 2D MDS projection of the sensory symbols when a sawtooth function links together world gray values to sensory symbols. Each symbol is represented as a circle whose color represents the *internal* coding. The corresponding symbols *in the outside world* are represented as a looping arrow around the projection. *Internal* black (symbol 0) and white (symbol 255) symbols are now close to each other, differently from Figure 1g.

continuity initially captured in the previous experiment leads 546  
 to a looping representation where the two opposite symbols, 547  
 0 and 254, are now close to each other in the internal rep- 548  
 resentation as they both correspond to close grayscale values 549  
 in the environment. Such a conclusion might be obvious in 550

551 this specific case, but it highlights that the *internal, subjective*  
 552 representation of the sensory symbols' topology might actually  
 553 differ greatly from our initial intuition as it depends on the  
 554 way the agent's sensors encode sensory information. The same  
 555 remark could apply to faulty sensors, whose output symbols  
 556 could be modified or rearranged because of some failure in  
 557 the information acquisition process; the proposed approach  
 558 could then allow the agent to (re)build an adequate internal  
 559 representation, though still intrinsically limited by its own  
 560 defective sensory capabilities.

### 561 C. Results for color perception

562 To further illustrate the approach, we will now endow the  
 563 agent with some color perception capabilities. Then, in this  
 564 subsection, the initial color tuples  $(R_{ij}, G_{ij}, B_{ij}) \in \llbracket 0; 255 \rrbracket^3$   
 565 coding the video pixel values  $v_{ij}$  are now mapped to the  
 566  $S = \alpha^3$  agent's sensels values  $s_{ij} \in \llbracket 0; \alpha^3 - 1 \rrbracket$  along

$$s_{ij} = g(v_{ij}) = Q_\alpha(B_{ij}) + \alpha Q_\alpha(G_{ij}) + \alpha^2 Q_\alpha(R_{ij}), \quad (18)$$

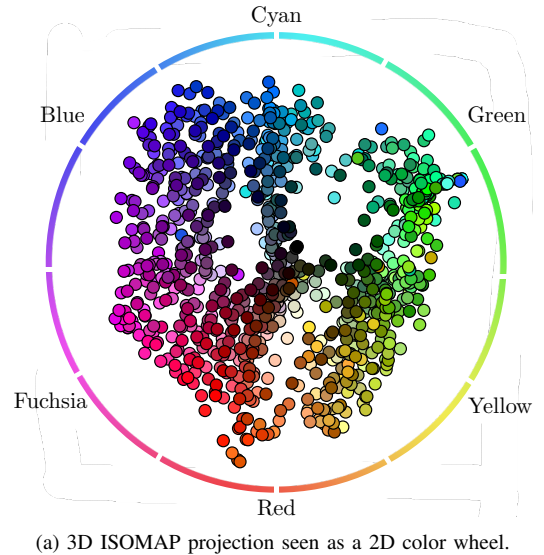
567 with  $Q_\alpha(\cdot)$  a quantification function defined by

$$Q_\alpha : X \mapsto Q_\alpha(X) = \text{round} \left( \frac{X}{255} \times (\alpha - 1) \right), \quad (19)$$

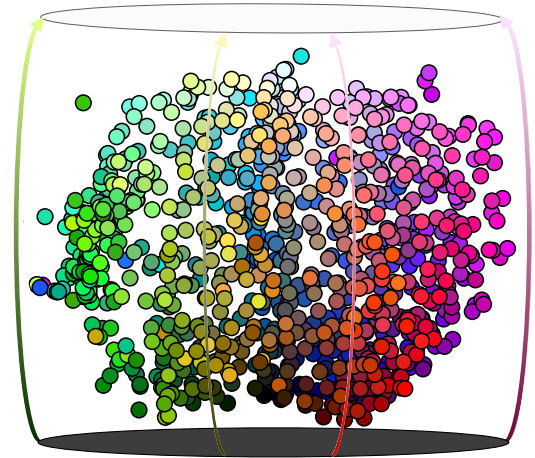
568 with  $X \in \llbracket 0; 255 \rrbracket$  and  $Q_\alpha(X) \in \llbracket 0; \alpha - 1 \rrbracket$ . Note that  
 569 while the symbol ordering was quite obvious for grayscale  
 570 values from an external point of view (e.g., the natural order  
 571 from 0 to 255) for the various matrices  $M$ ,  $P$ ,  $\Delta$ , and  $D$ ,  
 572 this no longer holds for these color sensory output symbols.  
 573 Nevertheless, the order in which they appear as line or column  
 574 indices in these matrices is not relevant since the only relevant  
 575 information about their closeness is entirely independent of  
 576 how these symbols are actually ordered. In all the following,  
 577  $\alpha = 10$  is selected, so that the agent's sensory space is made  
 578 of  $S = \alpha^3 = 1000$  uninterpreted (numerical) symbols. On  
 579 this basis, all the previous steps are successively applied. The  
 580 resulting  $D$  matrix can then be visualized through a low-  
 581 dimensional embedding technique like ISOMAP [28]. The  
 582 result of this projection performed in 3D is shown in Figure 3.  
 583 The obtained representation is in line with some classical  
 584 representations of RGB color models, like the HSL or HSV  
 585 coding of color. Indeed, the 3D point cloud first appears to  
 586 capture some color order very similar to the classical hue  
 587 color wheel, where pure colors are represented through an  
 588 angular position on a circle, as depicted in Figure 3a. But the  
 589 3D projection also exhibits a third axis linking very dark to  
 590 very light shades for each color of the hue wheel, similar to  
 591 the lightness axis in the HSV color coding, see Figure 3b. In  
 592 order to assess in a more quantitative way the similarity of  
 593 this low-dimensional projection with different color models, it  
 594 is proposed to compute a Frobenius distance  $\mathcal{F}$  along

$$\mathcal{F} = \left\| |D| - |D_m| \right\|, \quad (20)$$

595 where  $D_m$  is the  $S \times S$  distance matrix between all the sensory  
 596 symbols observed during the experiment for the color model  
 597  $m \in \{\text{RAND}, \text{RGB}, \text{HSL}, \text{HSV}\}$ , and  $|\cdot|$  the standardization  
 598 operator. The resulting distances are reported in Table I,  
 599 where the HSV color model better fits the obtained projection,  
 600 as initially qualitatively intuited. The same study can be  
 601 conducted by computing the k-nearest-neighbors between the



(a) 3D ISOMAP projection seen as a 2D color wheel.



(b) The same 3D projection seen as a cylinder, with the lightness axis drawn as arrows from black to white.

Figure 3: Interpretation of the 3D ISOMAP projection of the matrix  $D$  when the agent is endowed with color perception capabilities. (a) Representation obtained when viewing the projection “from below”: we can notice that all the sensory symbols are arranged by color, matching the intuitive color wheel which has been added to the graph. (b) Another point of view on the 3D sensory symbols representation: in addition to the color ordering highlighted in subfigure (a), a third axis supports the variation of lightness. The obtained projection can thus be understood as analogous to the HSL cylinder or biconic representation of the RGB color model.

obtained representation and the different color models. The  
 result of such a study is also reported in Table I for  $k = 75$   
 and exhibits the same conclusion. But one still has to keep in  
 mind that finding the best fitting color model is not important  
 by itself, since it is only exploited to *illustrate* the smooth  
 transitions from one color symbol to another, without apparent  
 discontinuity in the low dimensional embedding, as a way to  
 represent the information actually captured by the agent in  $D$ ,  
 which is, in the end, the only data it exploits in the following.  
 Then, with such a representation, the agent is now able to  
 assess if the sensory symbol associated with the rose color is  
*closer* to the one associated with the red color than it is to the  
 green one thanks to its internal metric matrix  $D$ .



Color model	RAND	RGB	HSL	HSV
Distance $\mathcal{F}$	1320	947	890.4	857.2
75-NN rate	8.4%	44.4%	45.5%	47.5%

Table I: Comparison between different color models, with RAND representing a random organization of the observed color symbols. (2nd line) Distances  $\mathcal{F}$  between the low-dimensional projection and the corresponding color model. (3rd line) Rate of 75 nearest neighbors between the obtained representation and the corresponding color model.

#### IV. LOCOMOTIVE MOTIVES: A CASE FOR FITTING EXPLORATION AND EXPLOITATION DYNAMICS

The previous developments were largely devoted to the relationship between two internal observations: the transition probabilities and the resulting metric. We showed how this information allows the agent to build some notion of closeness between sensory symbols –which could be understood as some subjective notion of sensory continuity– from certain successions of sensory experiences being more likely, or *typical*, than others. But such considerations clearly rely on the idea that typical environment states also display certain typical patterns themselves. From an external point of view, one would certainly declare that “environment states are (mostly) continuous”, both in time and space. This underlying assumption has not been dealt with so far, especially since the agent was passively observing sensory symbols changing over time in the previous experiments and not actively exploring its environment. This section thus aims to study which external structures in the states of the environment could explain the relationships between the agent’s motor actions associated with a sensory experience and the observed regularity, effectively giving action a defining role in this internal assessment. We then propose to study the influence of the agent’s action amplitude on its subjective sensory symbol continuity when it interacts with a mostly continuous environment. To that end, additional formal considerations are introduced in the first subsection. On this basis, some new experiments are proposed in the second subsection to highlight the importance of movement in building this subjective continuity.

##### A. Fitting spatial and sensory dynamics in the exploration

1) *Spatial and temporal coherence*: The results obtained in Section III were based on a purely passive observation of a changing “natural” visual scene –where the word *natural* here refers to our own usual and intuitive sensorimotor experience– allowing the agent to build a metric on its sensory symbols. But this distance should highly depend on the environment states and the successive configurations with which the agent samples it along time. More precisely, this implies that the environment’s state should exhibit some typical patterns, both in space and time, in line with the manner in which the agent conducts its interaction, to make apparent the notion of certain sensory symbol transitions being “more typical” than others. Thus, one first condition to fulfill is *spatial*, mandating that e.g. immediately next to a red region  $\mathcal{X}'$  of ambient space  $\mathcal{X}$  it is more likely to be another region  $\mathcal{X}''$  that is orange than cyan itself. In other words, we would generally expect the two events

$$\{\gamma_\epsilon(t)|_{\mathcal{X}'} = \epsilon_0\} \text{ and } \{\gamma_\epsilon(t)|_{\mathcal{X}''} = \epsilon_1\} \quad (21)$$

to largely depend on one another when  $\mathcal{X}'$  and  $\mathcal{X}''$  denote close (and small) regions of space. Furthermore, one second condition is *temporal*, so that the environment’s state at any localization  $\mathcal{X}'$  does not immediately change too randomly, so that the two events

$$\{\gamma_\epsilon(t)|_{\mathcal{X}'} = \epsilon_0\} \text{ and } \{\gamma_\epsilon(t + \Delta t)|_{\mathcal{X}'} = \epsilon_1\} \quad (22)$$

are conditioned on one another when  $\Delta t$  remains sufficiently small. We should insist on the fact that this coherence property, however, should only be local and relative to the agent’s exploration dynamics. It is clear that the color of a point  $x \in \mathcal{X}$  and time  $t \in \mathcal{T}$  does not depend on which colors appear two kilometers away, one and a half days from there. On the other hand, should the agent instead perform a two-kilometers long movement between two successive time samples, it should not be able to infer any relationship between successive sensory readings from the sole spatial coherence constraints.

2) *A formal account of spatiotemporal coherence*: Let us now generalize the previous sensory transition probabilities (6) by introducing, for any (sub)collection of motor trajectories  $\mathcal{B}'_{\mathcal{T}} \subseteq \mathcal{B}_{\mathcal{T}}$ ,

$$P_{s'|s}^{\mathcal{B}'_{\mathcal{T}}} = \{\gamma_s(t+1) = s' \mid \gamma_s(t) = s \text{ and } \gamma_{\mathbf{b}} \in \mathcal{B}'_{\mathcal{T}}\}, \quad (23)$$

for which  $P_{s'|s}^{\mathcal{B}'_{\mathcal{T}} = \mathcal{B}_{\mathcal{T}}} = P_{s'|s}$ . Such a (slight) generalization allows to highlight how a specific set of motor trajectories actually condition the sensory transitions available in the agent’s sensorimotor flow. More precisely, we introduced in [16] the *sensor receptive field* as the specific region of space for which the state of the environment is sufficient to fully determine the agent’s sensory state  $\mathbf{s}$ . Formally, a sensor receptive field can be seen as a function  $F : \mathbf{b} \in \mathcal{B} \mapsto F(\mathbf{b}) \subset \mathcal{X}$  verifying

$$\forall \epsilon_1, \epsilon_2 \in \mathcal{E}, \forall \mathbf{b} \in \mathcal{B}, \quad \epsilon_1|_{F(\mathbf{b})} = \epsilon_2|_{F(\mathbf{b})} \Rightarrow \psi(\mathbf{b}, \epsilon_1) = \psi(\mathbf{b}, \epsilon_2) = \mathbf{s}. \quad (24)$$

Then, let us now consider  $\mathcal{B}'_{\mathcal{T}}$  as a set of motor explorations  $\gamma_{\mathbf{b}}$  such that the receptive fields  $F(\gamma_{\mathbf{b}}(t))$  and  $F(\gamma_{\mathbf{b}}(t+1))$ , which condition successive sensory outputs  $\gamma_s(t)$  and  $\gamma_s(t+1)$ , fall *far apart* from one another. Then, based on our prior assumptions, the corresponding local environment states  $\gamma_{\epsilon|F(\gamma_{\mathbf{b}}(t+1))}(t+1)$  and  $\gamma_{\epsilon|F(\gamma_{\mathbf{b}}(t))}(t)$  should be independent: the physical properties available to the agent in the environment, restricted to the regions of space it would sample at time  $t$  and  $t+1$  by following a motor trajectory  $\gamma_{\mathbf{b}} \in \mathcal{B}'_{\mathcal{T}}$ , should not depend on each other. It then follows that  $\gamma_s(t+1) = \gamma_{\gamma_{\mathbf{b}}(t+1), \epsilon|F(\gamma_{\mathbf{b}}(t+1))}(t+1)$  and  $\gamma_s(t) = \gamma_{\gamma_{\mathbf{b}}(t), \epsilon|F(\gamma_{\mathbf{b}}(t))}(t)$  should be independent themselves. As a result, we have

$$P_{s'|s}^{\mathcal{B}'_{\mathcal{T}}} = \{\gamma_s(t+1) = s' \mid \gamma_s(t) = s, \gamma_{\mathbf{b}} \in \mathcal{B}'_{\mathcal{T}}\}, \quad (25)$$

$$= \{\gamma_s(t+1) = s' \mid \gamma_{\mathbf{b}} \in \mathcal{B}'_{\mathcal{T}}\}.$$

Thus, the probability  $P_{s'|s}^{\mathcal{B}'_{\mathcal{T}}}$  –where  $\mathcal{B}'_{\mathcal{T}}$  is a motor trajectory for which the coherence properties mentioned before are not verified– should not depend on previous sensory output  $\mathbf{s}$  anymore; instead, it simply replicates the *unconditional* probability that the agent experiences the particular sensory value  $s'$ . To the agent, this means that the knowledge of which information  $\mathbf{s}$  it experiences at timestep  $t$  does not give it *any* information on which sensation  $s'$  it is poised to experience at

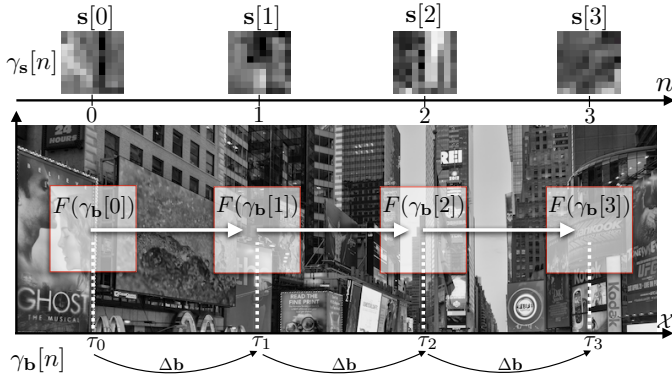


Figure 4: Experimental setup to assess the effect of the agent’s movement on the internal sensory symbol topology. A camera, whose field of view –or receptive field– is drawn as a square with red borders, faces a grayscale image and moves from one position  $\tau$  to another thanks to an action of amplitude  $\Delta\mathbf{b}$ . The corresponding sensory states  $\mathbf{s}[n]$  are then captured along time to build the statistics of the sensory symbol transitions.

713  $t + 1$ . Importantly, this shows that suitable choices of motor  
714 explorations are required for building a valid sensory metric,  
715 as well as giving an internal observation to assess whether  
716 this condition fails through Equation (25). The influence of  
717 this motor exploration is experimentally studied in the next  
718 subsection to illustrate these developments.

#### 719 B. An experimental assessment of the influence of the move- 720 ment amplitude

721 We propose in this subsection to assess the effect movement  
722 of the agent has on the internal representation of sensory  
723 symbols through simple simulations, where the agent is now  
724 allowed to move in a fixed environment. To begin with, details  
725 about the experiment setup are given. Next, the resulting  
726 representations are analyzed and discussed.

727 1) *Experimental setup*: let us consider in all the following  
728 a very simple agent, whose body is made of a planar, rectan-  
729 gular, camera sat atop one actuator, allowing the agent to only  
730 move in one direction, see Figure 4. The pixels of the camera  
731 are sensitive to the luminance of the ambient stimulus, which is  
732 a fixed grayscale image placed in front of the moving camera.  
733 In such a case, the ambient space  $\mathcal{X}$  is then the plane  $\mathbb{R}^2$ , and  
734 the state of the environment is a function  $\epsilon$  mapping a position  
735  $(x, y)$  in the plane to luminance values  $\epsilon(x, y) \in \llbracket 0; 255 \rrbracket$   
736 as encoded in the grayscale image. Those values are then  
737 converted into a sensory vector  $\mathbf{s} \in \llbracket 0; 255 \rrbracket^{W \times H}$  directly  
738 capturing the corresponding grayscale value in the environ-  
739 ment (the function  $h(\cdot)$  in Equation (16) is thus the identity  
740 function). In the forthcoming simulations,  $W = H = 100$ .  
741 As already outlined in §II-A, we consider the agent is able to  
742 move in its environment by applying a single *action*  $a$  [16], i.e.  
743 by applying a function  $a$  to its current absolute configuration  
744  $\mathbf{b} = (\mathbf{m}, \tau)$  to go to another configuration  $\mathbf{b}' = (\mathbf{m}', \tau') = a\mathbf{b}$ .  
745 In this section, we will mainly study the influence of the  
746 amplitude  $\Delta\mathbf{b}$  of this action, which is supposed to produce  
747 a movement of the camera in only one direction and with the  
748 same amplitude, as illustrated in Figure 4. This is obviously  
749 a very particular and restrictive action, at least in comparison  
750 with the more generic motor action framework presented by

the authors in [16], but it will still allow a comprehensive  
study of the effect of movement on the internal representation  
built by the agent. The different action amplitudes  $\Delta\mathbf{b}$  used  
in the simulations will all be equal to a multiple of  $\sigma_{\mathbf{b}}$ , a  
particular amplitude which causes a shift of the perceived  
information in  $\mathbf{s}$  of exactly 1 pixel. This actually corresponds  
to a displacement of the camera receptive field  $F(\mathbf{b})$  in  $\mathcal{X}$   
(represented as squares with red borders in Figure 4) of the  
width of 1 pixel in the plane supporting the grayscale image.  
Note that the amplitude of the actions is explicitly an *external*  
metric that is not available to the agent; all it knows about  
is that it is using an action with an unknown amplitude to  
move in its environment. We will show in §IV-B3 that, under  
the proposed assumptions, the agent will be able to compare  
the amplitude of its actions on the basis on their sensory  
consequences.

In practice, the experiment is conducted the following  
way: to begin with, the environment observed by the agent  
is a grayscale image of a crowded street, partially shown  
in Figure 4. Then, starting from a fixed (random) position  
 $\tau_0$  in the environment, the agent follows a motor trajectory  
 $\gamma_{\mathbf{b}}[n]$  made of jumps of fixed amplitude  $\Delta\mathbf{b}$ . This produces  
a displacement of the agent’s sensor receptive field in the  
environment at which the agent gathers samples  $\mathbf{s}[n]$  of its  
corresponding sensory trajectory  $\gamma_{\mathbf{s}}[n]$ . After having generated  
 $N_a$  times the same action  $a$ , the camera is put in one other  
random position in the image; then, the action  $a$  is used again  
to move the camera  $N_a$  times in the image. This process is  
repeated  $N_r$  times, so that  $N_r \times N_a$  sensory samples  $\mathbf{s}[n]$  are  
collected. These samples then allow one to build the matrix  $P$   
as in Equation (13). Then, the corresponding MDS projection  
of the distance matrix  $D$  can be computed to visualize the  
captured sensory symbol topology. The experience is finally  
repeated for various amplitudes  $\Delta\mathbf{b}$ .

2) *Results*: The experiment has been conducted for  $N_a =$   
500 and  $N_r = 100$ , so that  $50.10^3$  sensory transitions are  
used to build the matrix  $P$  for each action amplitude  $\Delta\mathbf{b}$   
chosen among  $\{\sigma_{\mathbf{b}}, 25\sigma_{\mathbf{b}}, 250\sigma_{\mathbf{b}}, 1000\sigma_{\mathbf{b}}\}$ . Note that being  
greater than the size  $N_s = 100$  of the sensor,  $\Delta\mathbf{b} = 250\sigma_{\mathbf{b}}$   
leads to the sampling of an area in the environment that does  
not overlap with its previous receptive field position. Figure 5  
represents successively (in each row) the matrices  $P, D$  and  
their corresponding embedding  $\text{MDS}_2(D)$  for each of the 4  
selected amplitudes (in each column). Let us first consider  
the evolution of the probability matrix as a function of the  
movement amplitude (first row). For a very small action ampli-  
tude, the probability matrix  $P$  exhibits a clear diagonal pattern  
indicating that close sensory symbols (in terms of gray levels)  
correspond to high transition probabilities; qualitatively, we  
face more or less the same conditions than in §III-B1 where the  
observation of the environment changes matches the changes  
in perception induced by the action of the agent. These two  
scenarios no longer correspond when the action amplitude  
rises: the higher the amplitude, the wider the probability  
distribution. For the largest amplitude, the diagonal pattern  
cannot even be seen anymore in  $P$  and high probabilities do  
not correspond to close gray levels anymore. This tendency  
is clearly confirmed when computing the distance matrix  $D$   
from  $P$  (2nd row in Figure 5): for high amplitudes, mostly all

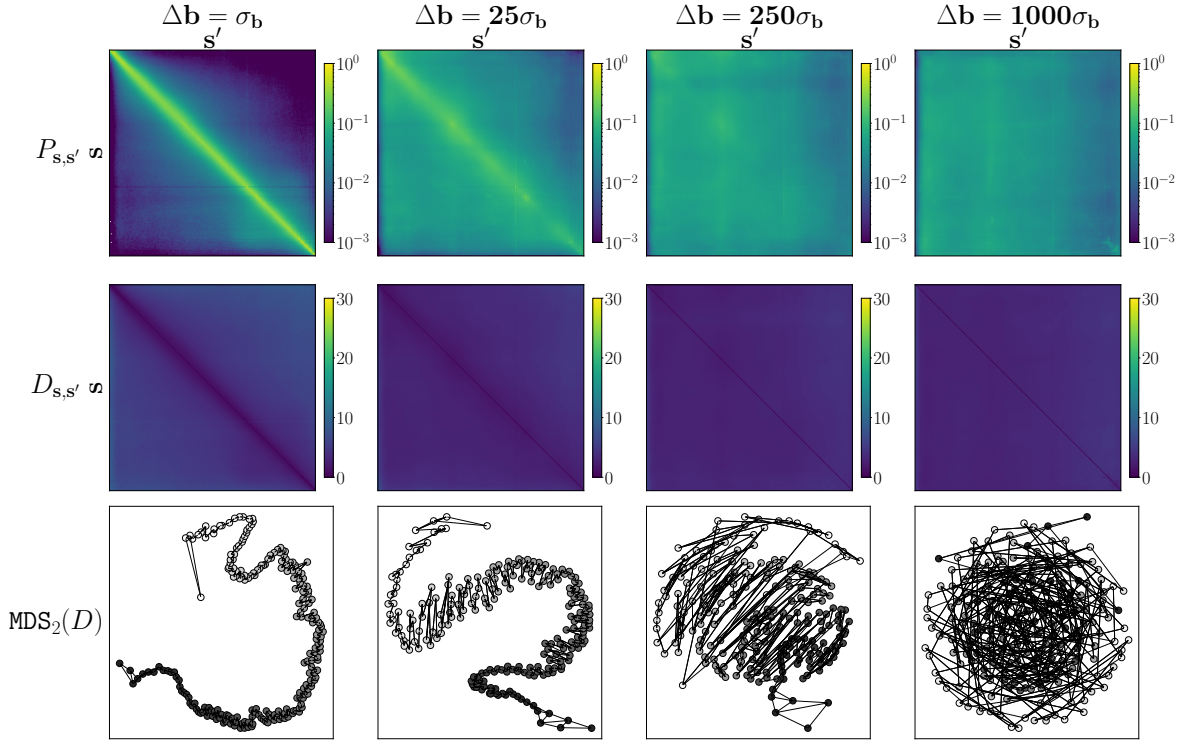


Figure 5: Evolution of the transition probability measure  $P$  (displayed with a logarithmic norm), the distance measure  $D$  and the representation of this distance projected using 2-dimensional MDS for increasing movement amplitudes. Each column represents a motor trajectory for a fixed amplitude  $\Delta \mathbf{b}$  described relative to a 1 pixel shift of its sensor’s field of view. We can see that the diagonal pattern for  $P$  and  $D$  as well as the uni-dimensional grayscale manifold are deteriorating as the movement amplitude gets bigger, indicating the inability to capture spatial coherence properties. The links that connect each symbol on the MDS representation are a  $k$ -NN like algorithm that assess how the agent perceive its symbol continuity.

810 symbols are now close to each other. Obviously, this results in  
811 very different  $2D$  projections of the matrix  $D$  (3rd row). For  
812 the lowest amplitudes, we still clearly see a one-dimensional  
813 manifold, folding on itself when the action amplitude grows.  
814 But the dimensionality of the manifold is not sufficient to  
815 tell if the agent correctly captured or not the sensory symbol  
816 topology. Like we did in Figure 1g and Figure 1d, a  $k$ -NN  
817 algorithm is computed on  $D$  and displayed in Figure 5 to link  
818 each symbol to its closest neighbor, this link being represented  
819 as a line between two symbols in the projection. Looking  
820 at the smallest amplitude, the sensory symbols manifold can  
821 be browsed in the usual grayscale order by following the  
822 aforementioned lines. On the contrary, this proves impossible  
823 for larger amplitudes, where lines link together e.g., symbols  
824 associated with clear and dark gray levels. It is then clear  
825 that the conditions written as Equations (21) and (22) are not  
826 verified anymore, with two successively sampled environment  
827 states associated with two distant positions in space, leading to  
828 the loss of perception by the agent of the spatial and temporal  
829 coherence in the environment.

830 3) *Discussions:* Equation (25) states that, for a specific set  
831 of motor trajectories  $\mathcal{B}'_{\mathcal{T}}$  making successive receptive fields  
832 falling apart from one another, the probability of transition  
833 between successive sensory symbols tends to an *unconditional*  
834 probability  $P_{s'|s}^{\mathcal{B}'_{\mathcal{T}}} = P_{s'}$ . Importantly, this phenomenon can  
835 be *internally* assessed by the agent since both probability  
836 distributions are only based on sensory symbol observations;

this then constitutes some *internal* way for the agent to rate 837  
the spatial and temporal coherence of its interaction with 838  
the environment. To that end, we propose to compare the 839  
two probability distributions  $P_{s'|s=s_k}$  –the probability of every 840  
sensory value to succeed to a specific sensory value  $s_k$ – 841  
and  $P_{s'}$ , by using the Jensen-Shannon distance  $D_{JS}$  [29], a 842  
bounded metric based on the symmetrized version of the 843  
Kullback–Leibler divergence [30], and defined as 844

$$D_{JS}(P_{s'|s=s_k} \| P_{s'}) = \sqrt{\frac{\text{KL}(P_{s'|s=s_k} \| M) + \text{KL}(P_{s'} \| M)}{2}}, \quad 845$$

$$\text{with } M = \frac{1}{2} (P_{s'|s=s_k} + P_{s'}), \quad 846$$

with the KL divergence for two discrete probabilistic distribu- 848  
tions  $A$  and  $B$  defined in the probability space  $\mathcal{W}$  as 849

$$\text{KL}(A \| B) = \sum_{x \in \mathcal{W}} A(x) \log_2 \left( \frac{A(x)}{B(x)} \right). \quad 850$$

This results in a distance  $D_{JS}(\cdot)$  between 0 and 1, computed 852  
for each sensory symbol  $s'$ , that is expected to converge 853  
towards 0 when both distributions are identical, i.e. when the 854  
motor trajectory of the agent leads to having  $s$  and  $s'$  inde- 855  
pendent. Again,  $D_{JS}$  is computed for 4 different amplitudes 856  
 $\{\sigma_b, 5\sigma_b, 25\sigma_b, 125\sigma_b\}$  with corresponding graphs in Figure 6. 857  
The results displayed in Figure 6 show that the JS distance for 858  
every probability distribution systematically decreases when 859  
the amplitude of the agent’s movement increases, i.e. when 860

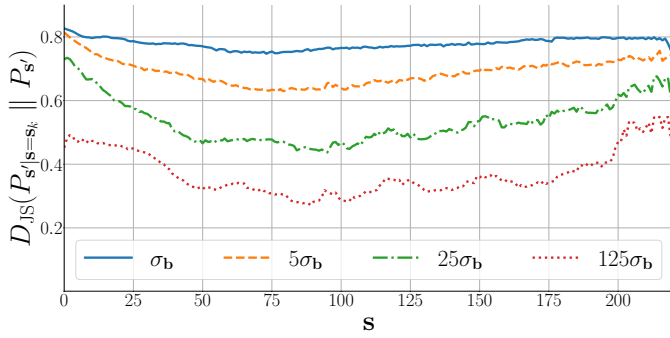


Figure 6: JS distance of every conditional probabilities relatively to the unconditional probability  $P(s')$  for different movement amplitudes. Each point of the plot represents a JS distance for a single conditional probability to  $P(s')$ . As the amplitude increases, the divergence of every symbol decreases, getting close to the unconditional probability.

$\Delta \mathbf{b}$	$\sigma_{\mathbf{b}}$	$5 \sigma_{\mathbf{b}}$	$25 \sigma_{\mathbf{b}}$	$125 \sigma_{\mathbf{b}}$
$I(\mathbf{s}, \mathbf{s}')$	5.36	3.49	1.92	0.94

Table II: Mutual information between sensory symbols  $\mathbf{s}$  and  $\mathbf{s}'$  as a function of the agent's action amplitude.

the conditional distribution tends towards the unconditional one as described by Eq. (25). In the same vein, we can also conduct this comparison by computing the mutual information between, roughly speaking, the sensory symbols before and after the agent's movement and defined by

$$I(\mathbf{s}, \mathbf{s}') = \sum_{k,l} p_{s_k, s_l} \log_2 \left( \frac{p_{s_k, s_l}}{p_{s_k} p_{s_l}} \right), \quad (28)$$

with  $P_{\mathbf{s}, \mathbf{s}'} = (p_{s_k, s_l})_{s_k, s_l}$  the joint probability and  $P_{\mathbf{s}} = (p_{s_k})_{s_k}$  and  $P_{\mathbf{s}'} = (p_{s_l})_{s_l}$  the marginal probabilities. This mutual information is computed for the same 4 amplitudes as in Figure 6, and is reported in Table II. As expected, the mutual information drops significantly by about 64% as soon as the movement amplitude rises to  $25\sigma_{\mathbf{b}}$ , showing again how the link between  $\mathbf{s}$  and  $\mathbf{s}'$  is degraded when the agent's motion amplitude becomes higher between two time steps. Importantly, these two comparisons between the two probability distributions could provide the agent with an *internal* way to rate the adequation of its motor exploration performed by applying actions with (at least for now) unknown consequences, or even an internal signature of the amplitude of its own actions.

## V. USING THE METRIC TO GET AN INTERNAL ASSESSMENT OF SENSORY REGULARITY

Now that we have been able to quantify how and why the agent's action modulates its sensory symbol topology, let us focus on a more experimental use of the obtained representation. Intuitively, and thanks to the introduction of the metric  $d_f$ , the agent should now be able to assess if a sensory transition is typical or not. This could be used as a way to deal with the presence of noise in the raw sensory data, i.e. by being able to discriminate close (but not strictly equal) sensory values from irregular sensory transitions due to the presence of specific events in the environment (movement of an object in the scene, changes in the illumination conditions, etc.). This

section thus aims to present how the agent could internally assess its sensory regularity by first depicting some simple formal elements in the first subsection. Then, the second subsection shows how a naive agent could actually be capable of performing a sensory prediction task, even in the presence of noise, in the vein of the sensorimotor action framework presented by the authors in [16].

### A. Internally rating the sensory regularity

1) *Some formal considerations:* to begin with, let us consider again Eq. (7) by which  $\delta_f(\mathbf{s}, \mathbf{s}')$  is defined in terms of the sensory transition probabilities  $P_{s'|s}$ . It can be trivially rewritten as

$$\forall \mathbf{s}, \mathbf{s}' \in \mathcal{S}, \mathbb{P}(\gamma_s(t+1) = \mathbf{s}' \mid \gamma_s(t) = \mathbf{s}) = f^{-1}(\delta_f(\mathbf{s}, \mathbf{s}')), \quad (29)$$

when  $f$  is injective. But because  $f$  is also necessarily non-increasing, so must be  $f^{-1}$ ; this obviously entails that the probability of any sensory transition from  $\mathbf{s}$  to  $\mathbf{s}'$  is as expected a decreasing function of the sensory distance between them. However, we also know from the definition of the metric  $d_f$  from shortest paths in Eq. (9) that

$$\forall \mathbf{s}, \mathbf{s}' \in \mathcal{S}, d_f(\mathbf{s}, \mathbf{s}') \leq \delta_f(\mathbf{s}, \mathbf{s}'). \quad (30)$$

We then have immediately

$$\forall \mathbf{s}, \mathbf{s}' \in \mathcal{S}, \mathbb{P}(\gamma_s(t+1) = \mathbf{s}' \mid \gamma_s(t) = \mathbf{s}) \leq f^{-1}(d_f(\mathbf{s}, \mathbf{s}')). \quad (31)$$

Then, Eq. (31) guarantees that, from any sensory value  $\mathbf{s}$ , the probability to land on  $\mathbf{s}'$  at a distance  $d_f(\mathbf{s}, \mathbf{s}') = \lambda$  is therefore *less than*  $f^{-1}(\lambda)$ . This property thus gives an intrinsic way of quantifying the *regularity* of a transition in the sensory experience. Indeed, providing some “metric rejection threshold”  $\tau_r$ , the agent might be able to deem all sensory transitions  $\mathbf{s}$  to  $\mathbf{s}'$  of corresponding distance  $d_f(\mathbf{s}, \mathbf{s}')$  as irregular (resp. regular) whenever  $d_f(\mathbf{s}, \mathbf{s}') \geq \tau_r$  (resp.  $d_f(\mathbf{s}, \mathbf{s}') < \tau_r$ ). Obviously, determining whether a transition is regular or not might also be decided directly on the basis of its probability of transition. To stay consistent, we choose to only investigate the properties of our regularity measure from the standpoint of the sensory metric.

Still, one should notice that Eq. (31) is merely an inequality, as opposed to the corresponding equality in Eq. (29). To the agent, this means that there may be some particular transitions from  $\mathbf{s}$  to  $\mathbf{s}'$  which are still unlikely even if the agent found  $d_f(\mathbf{s}, \mathbf{s}')$  to be small: basically, this criterion can allow *false positives*, while it guarantees that all transitions rejected on the basis of this metric verify the occurrence probability inequality, that is it does not cause *false negatives*.

2) *Example:* we selected in previous sections (see Eq. (14)) the function  $f = -\log$  to map the estimated transition probabilities  $p_{kl}$  to the metric prototype  $\delta_{kl}$ . In such a case, still with a threshold  $\tau_r$ , an irregular transition should then typically occur with a probability  $\mathbb{P}(\gamma_s(t+1) = \mathbf{s}' \mid \gamma_s(t) = \mathbf{s}) \leq e^{-\tau_r}$ . Then, selecting for instance the threshold values  $\tau_r \in \{1, 3, 5\}$  will allow the agent to reject transitions that occur in less than about  $\{37, 5, 1\}\%$  of occurrences.

## 940 B. Exploiting the sensory regularity for sensory prediction

941 We now propose to exploit the agent’s capability to decide  
 942 whether a sensory transition is regular in a sensory prediction  
 943 task in the presence of noise inside the sensory data. To that  
 944 end, the framing of the approach in [16] is first briefly intro-  
 945 duced, followed by the proposed experimental setup, mirroring  
 946 that of this previous contribution. Then, the sensory prediction  
 947 framework from [16] is applied for different scenarios: (i) with  
 948 no noise in the sensory data or with noise, but (ii) without or  
 949 (iii) with rating the sensory regularity. A discussion comparing  
 950 these scenarios is then proposed in the second paragraph.

951 1) *A short recall on the framing of the problem:* The  
 952 contribution from [16] is all about the theoretical conditions  
 953 for the determination of a sensory prediction function for  
 954 a naive agent. More precisely, it is demonstrated how the  
 955 algebraic structure found in this prediction is homeomorphic  
 956 to that of an algebraic group of specific motor actions, the  
 957 *conservative* actions. An action  $a$  is said to be conservative  
 958 if all sensels of the agent exchange the places they sample  
 959 when applying  $a$ : equivalently, conservative actions can then  
 960 be thought of as permutations of sensels. Importantly, this  
 961 result has since been extended to *quasiconservative* actions  
 962 in [23], where partial sensory prediction maps are proposed  
 963 to generalize the sensel permutations of strictly conservative  
 964 actions for the case where some sensels have no identified  
 965 permutations when applying an action  $a$  (e.g. for sensels in  
 966 the border of a camera).

967 2) *Experimental setup:* the proposed simulation setup is  
 968 very close to the one already presented in §IV-B1. The agent  
 969 is still made of a moving camera facing a fixed grayscale  
 970 image, as shown in Figure 4. This time, the agent is endowed  
 971 with a  $W \times H = 10 \times 10$  sensor, and is now able to  
 972 move in four orthogonal directions by applying 5 different  
 973 (quasiconservative) actions:  $a_{id}$ ,  $a_f$ ,  $a_b$ ,  $a_r$  and  $a_l$  making the  
 974 camera receptive field respectively stay still, move in the left,  
 975 right, up or down directions in  $\mathcal{X}$ . It is clear that the sensory  
 976 consequences of such actions can be illustrated as a shift  
 977 of information in the image in the opposite direction of the  
 978 agent’s movement: most of the sensels values *before* applying  
 979 any of these actions can find a successor *after*. Then, predicting  
 980 the sensory consequence of an action can be summed up by a  
 981 permutation between sensel values, providing all sensels share  
 982 the same excitation function, as already outlined in §III-A2.  
 983 Importantly, the agent has no clue about the incidence of a  
 984 given action nor about their possible relationship. All it can  
 985 do is perform an action and observe its consequences in its  
 986 sensory data [16]. The proposed experimentation then relies  
 987 on the two following steps.

988 a) *Step 1: building of the sensory symbol topology.* The  
 989 agent explores its environment by repeatedly selecting random  
 990 actions in  $\mathcal{A} = \{a_{id}, a_f, a_b, a_r, a_l\}$  with identical amplitudes  
 991  $\Delta \mathbf{b} = \sigma_{\mathbf{b}}$  (apart from  $a_{id}$ ), and then infers the distance matrix  
 992  $D$ , in line with §IV-B where the number  $N_a$  of draws of  
 993 actions is set to  $N_a = 25$ , and the number of repetition is  
 994 selected to  $N_r = 2.10^3$ . As opposed to the previous case,  
 995 some artificial noise  $n_{ij}$  is now added to the pixel value  $v_{ij}$  of  
 996 the image to form the agent’s sensel values<sup>3</sup>  $s_{ij} = v_{ij} + n_{ij}$

<sup>3</sup>which is further clamped if need be, i.e. if  $s_{ij}$  exceeds 0 or 255, the sensel value is set to the closer bound.

before computing the matrix  $D$ , with  $n_{ij}$  a random integer  
 drawn from a centered discrete uniform distribution of width  
 $2\sigma_n$ .

b) *Step 2: building of the sensory prediction function.*  
 Once the matrix  $D$  is obtained, the agent performs a second  
 exploration of its environment so as to build a sensory predic-  
 tion function for each of its actions in  $\mathcal{A}$ . As previously argued,  
 these functions can take the form of binary permutation  
 matrices [16]  $\Pi_{a_p} = (\pi_{kl}^{(p)})_{k,l}$  of size  $N_s \times N_s$ , with  $a_p \in \mathcal{A}$   
 and  $N_s = W \times H$ , as each pixel value in the sensory array  
 is expected to shift in different positions depending on the  
 spatial effect of the performed action. In these matrices, having  
 $\pi_{kl}^{(p)} = 1$  indicates that the  $k^{\text{th}}$  sensel takes the value of  
 the  $l^{\text{th}}$  sensel after applying action  $a_p$ . For this experiment,  
 $N_a = 50.10^3$  and  $N_r = 1$ . Initially, every element  $\pi_{kl}^{(p)}$  of  
 the permutation matrices  $\Pi_{a_p}$  is initialized to 1, meaning that  
 all permutations between the agent’s sensels are possible for  
 action  $a_p$ . Then, each time this action is drawn from  $\mathcal{A}$ , the  
 agent can discard in  $\Pi_{a_p}$  some permutations by observing  
 that some sensel values do not switch with one another, then  
 updating the corresponding matrix elements to 0 as per the  
 update rule

$$\pi_{kl}^{(p)}[n+1] = \begin{cases} 1 & \text{iff } s_l[n] = s_k[n+1] \text{ and } \pi_{kl}^{(p)}[n] = 1 \\ 0 & \text{else,} \end{cases} \quad (32)$$

where  $s_k$  and  $s_l$  represent the sensel values associated with  
 the element at the position  $(k, l)$  in the permutation matrix  
 $\Pi_{a_p}$ . We can notice in Eq. (32) that the elements in these  
 matrices are set to 0 as soon as a permutation is not detected  
 by the strict equality between sensory values. This limitation,  
 already outlined in [16], makes this approach fall apart when  
 dealing with noise in the sensory data or when interacting  
 with a non-static environment. Benefiting from the previous  
 developments, we instead propose a revised update rule for  
 the permutation matrix as

$$\pi_{kl}^{(p)}[n+1] = \begin{cases} 1 & \text{iff } d_f(s_l[n], s_k[n+1]) < \tau_r \text{ and } \pi_{kl}^{(p)}[n] = 1 \\ 0 & \text{else,} \end{cases} \quad (33)$$

where  $\tau_r$  is a manually chosen threshold applied to the  
 built matrix distance  $D$ . In the following,  $\tau_r$  is tuned so  
 as to correspond to the smallest threshold that allows for  
 permutation matrices to converge. It is clear that this is a strong  
*a priori*, and the way the agent can autonomously set this  
 threshold is still an ongoing work, discussed in the conclusion.  
 In the same vein, the second step in this experiment requires  
 a second exploration of the *same* environment as during the  
 first step. This definitely a suboptimal process, only proposed  
 here to illustrate the benefits of the internal assessment of  
 sensory regularities for the proposed sensory prediction task.  
 Obviously, the sensory transitions observed when building the  
 sensory symbol topology could also be used for building the  
 sensory prediction functions. Importantly, this highlights again  
 the importance of the threshold  $\tau_r$  which should then also be  
 selected by an appropriate combination of these two steps.

c) *Evaluation: convergence of the permutation matrices.*  
 To evaluate the influence of the added noise on the conver-  
 gence of permutation matrices  $\Pi_{a_p}$ , we propose an (external)  
 criterion  $C(\Pi_{a_p}) = C_H(\Pi_{a_p}) \times C_D(\Pi_{a_p})$  adapted from [16]

1049 to account for the added noise to the data and defined along

$$C_D(\Pi_{a_p}) = \frac{\sum_{kl} \pi_{kl}^{(p)} \bar{\pi}_{kl}^{(p)}}{\sum_{kl} \bar{\pi}_{kl}^{(p)}}, \text{ and} \quad (34)$$

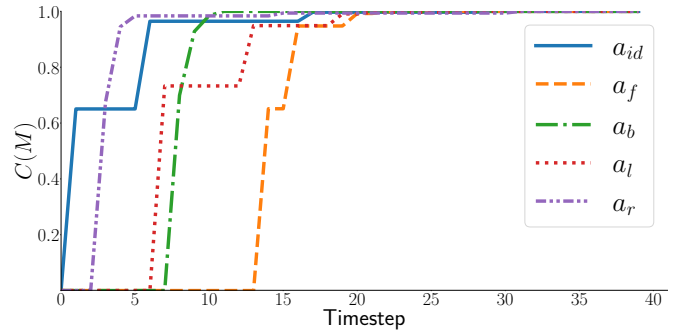
$$C_H(\Pi_{a_p}) = 1 - \frac{1}{N_s \log_2(N_s)} \sum_{i=1}^{N_s} H_i,$$

1050 with

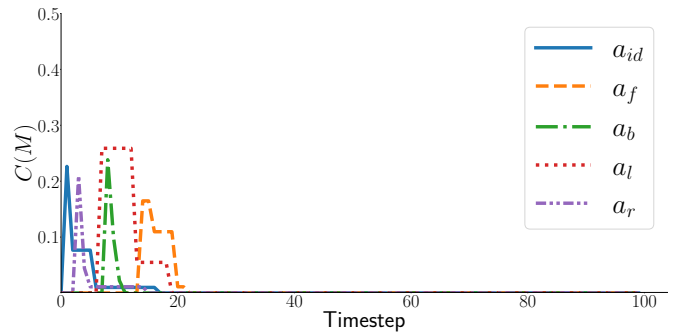
$$\begin{cases} H_i &= -\sum_{l=1}^{N_s} \frac{\pi_{kl}^{(p)}}{\mu_k} \log_2 \left( \frac{\pi_{kl}^{(p)}}{\mu_k} \right), \\ \mu_k &= \max \left( 1, \sum_{l=1}^{N_s} \pi_{kl}^{(p)} \right), \end{cases} \quad (35)$$

1051 where  $\bar{\pi}_{kl}^{(p)}$  represents the (binary) coefficients of the ideal  
1052 matrix  $\bar{\Pi}_{a_p}$  associated with the action  $a_p$ . Basically,  $C_H$  can be  
1053 understood as an average measure of certainty in the discovery  
1054 of the permutations, weighted by the percentage  $C_D$  of the  
1055 correctly identified permutations w.r.t. the ground truth to  
1056 account for the noise, possibly discarding some of them. In the  
1057 end, criterion  $C$  lies between 0 –i.e. the matrix is full of 1's  
1058 (initialization) or 0's (all permutations have been discarded)–  
1059 and 1 –i.e. the permutation has been correctly discovered.

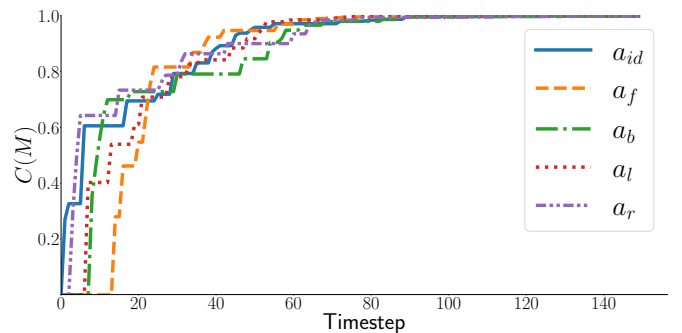
1060 3) *Results*: As outlined in the introduction of Section V,  
1061 three different scenarios are evaluated. To begin with, we  
1062 first consider the case where there is no noise in the agent's  
1063 perception by setting  $\sigma_n = 0$ . Then, using the update rule (32)  
1064 should allow the agent to correctly build all of its permutation  
1065 matrices, exactly as in [16]. As expected, Figure 7a shows  
1066 that criterion  $C$  converges towards its maximal value of 1 for  
1067 all actions in  $\mathcal{A}$ .  $C$  plots also exhibit sparse jumps at random  
1068 times, corresponding to the steps where the action was actually  
1069 drawn in  $\mathcal{A}$  during the experiment. More importantly, we can  
1070 see in Figure 7a that only a few realizations of each action  
1071  $a_p$  (about 4 to 6 here) are required for  $C(\Pi_{a_p})$  to almost  
1072 reach 1, showing how easy it is for the agent to discover the  
1073 existence of such permutations in its own perception. In the  
1074 second scenario, a noise of amplitude  $\sigma_n = 2$  is now added to  
1075 the sensation. Obviously, the strict comparison of sensel values  
1076 in (32) in the presence of such noise (however small) entirely  
1077 breaks the approach, as shown in Figure 7b. As expected, the  
1078 criterion  $C$  now converges to 0: each  $\Pi_{a_p}$  matrices converges  
1079 to null matrices as all possible permutations of values have  
1080 been (including erroneously) discarded in the process. Finally,  
1081 the new update rule (33) is now used to judge the closeness  
1082 of sensel values on the basis of the built distance  $D$ , resulting  
1083 in the evolution of the criterion  $C$  represented in Figure 7c.  
1084 For this scenario,  $\sigma_b = 1$  and  $\tau_r = 1.63$ . In the presence of  
1085 noise, the ability of the agent to assess if a sensation is now  
1086 close to others allows it to correctly discover the existence of  
1087 permutations in its perception. But clearly, this task is not as  
1088 easy as in the first scenario: the number of required actions for  
1089 correctly evaluating their corresponding permutation matrices  
1090 is significantly higher. This is apparent in Figure 7c, not only in  
1091 the slower convergence time of the criterion  $C$  but also in the  
1092 smaller jumps of values in  $C$ . Indeed, each generation of action  
1093 brings less information in the prediction process because of the  
1094 noise included in the agent's sensations. But still, the important  
1095 structures anchoring the sensorimotor interaction the agent has  
1096 with its environment are still available, allowing it e.g., to build  
1097 an image of its body [14] or of its peripersonal space [15], at



(a) Evaluation criterion  $C$  with  $\sigma_n = 0$  and strict equality update rule.



(b) Evaluation criterion  $C$  with  $\sigma_n = 2$  and strict equality update rule.



(c) Evaluation criterion  $C$  with  $\sigma_n = 2$  and a threshold in  $D$ .

Figure 7: Evolution of the evaluation criterion  $C$  for the 5 considered actions in  $\mathcal{A}$ . (a) With no noise and the update rule (32),  $C$  converges towards 1 in a very short number of realizations of each action. (b) In the same scenario, but with  $\sigma_n = 2$ , the update rule (32) do not allow to detect permutations anymore, resulting in the criterion falling down to 0. (c) When selecting a correct threshold  $\tau_r$  in Eq. (33), the agent is now able to build the 5 sensory prediction functions correctly, but with more realization of each action in comparison with (a).

1098 least at the cost of a longer interaction in time.

## 1099 VI. CONCLUSION

In this paper, and after purely topological considerations, 1100 a metric-based approach is proposed to formalize the ability 1101 of a naive agent to build some subjective sense of sensory 1102 continuity. An experimental framework is then proposed, il- 1103 lustrated and assessed in the context of visual perception for 1104 the discovery of gray or color scales. Then the importance of 1105 the dynamic of the agent's exploration relative to that of the 1106 environment is studied, highlighting an important spatiotem- 1107 poral coherence principle of this exploration. Finally, with a 1108 sensory closeness notion now available to the agent, a sensory 1109

prediction task is proved accessible even in the presence of noise, thus extending the robustness of this sensorimotor framework to realistic conditions.

Nevertheless, it is clear that this work still suffers from some limitations. For instance, the scalability of the proposed experimental framework is certainly limited. Indeed, although it was not the objective of this paper, the way the regularities are extracted from the raw sensations is certainly not computationally effective, considering the possibly very high number of sensory symbols involved in e.g., color perception for traditional camera sensors. Hierarchical approaches might be preferred [31], but still remain to be explored in the context of sensorimotor approaches to perception. Another limit concerns the notion of sensory neighbors: while being now formally accessible to the agent thanks to the proposed contribution, it still practically requires a threshold to be set w.r.t. the task to be performed. In this paper, this threshold has been manually tuned with two successive steps involving two independent explorations of the same environment, but we could instead rely on a closed-loop approach mixing the discovery of the sensory regularities with the corresponding sensory prediction task: as long as the prediction is not correctly built, the threshold must be adapted accordingly. Still, should the agent be able to perform some sensory prediction task, so should it be able to *quantitatively* compare its prediction with its actual perception. This should make it capable of detecting outliers in its environment, and in particular, changes in its perception that are not directly correlated to its own actions. This might be the way towards some internal notion of sensorimotor objects and thus would undoubtedly extend the scope of these approaches to more potential applications.

## REFERENCES

- [1] B. Dainton, "Temporal Consciousness," in *The Stanford Encyclopedia of Philosophy*, Winter 2018 ed., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2018.
- [2] J. M. B. Fugate, "Categorical perception for emotional faces," *Emotion Review*, vol. 5, no. 1, pp. 84–89, 2013.
- [3] J. C. Toscano, B. McMurray, J. Denhardt, and S. J. Luck, "Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech," *Psychological Science*, vol. 21, no. 10, pp. 1532–1540, 2010.
- [4] A. Herwig, "Transsaccadic integration and perceptual continuity," *Journal of Vision*, vol. 15, no. 16, pp. 7–7, 12 2015.
- [5] J. M. Stroud, "The fine structure of psychological time," *Annals of the New York Academy of Sciences*, vol. 138, no. 2, pp. 623–631, 1967.
- [6] R. VanRullen and C. Koch, "Is perception discrete or continuous?" *Trends in Cognitive Sciences*, vol. 7, no. 5, pp. 207–213, 2003.
- [7] J. O'Regan, *Why Red Doesn't Sound Like a Bell: Understanding the feel of consciousness*. Oxford University Press, 2011.
- [8] S. A. Morris, *Topology without tears*, 2020.
- [9] A. Censi, "Bootstrapping vehicles: A formal approach to unsupervised sensorimotor learning based on invariance," Ph.D. dissertation, California Institute of Technology, June 2012.
- [10] D. Philipona, J. K. O'Regan, and J.-P. Nadal, "Is there something out there?: Inferring space from sensorimotor dependencies," *Neural Comput.*, vol. 15, no. 9, pp. 2029–2049, 2003.
- [11] A. Laflaquière, J. K. O'Regan, S. Argentieri, B. Gas, and A. Terekhov, "Learning agents spatial configuration from sensorimotor invariants," *Robotics and Autonomous Systems*, vol. 71, pp. 49–59, September 2015.
- [12] A. Laflaquiere, "Grounding the experience of a visual field through sensorimotor contingencies," *Neurocomputing*, vol. 268, no. C, 2017.
- [13] A. V. Terekhov and J. K. O'Regan, "Space as an invention of active agents," *Frontiers in Robotics and AI*, vol. 3, p. 4, 2016. [Online]. Available: <https://www.frontiersin.org/article/10.3389/frobt.2016.00004>
- [14] V. Marcel, S. Argentieri, and B. Gas, "Building a sensorimotor representation of a naive agent's tactile space," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 2, pp. 141–152, June 2017.
- [15] V. Marcel, S. Argentieri, and B. Gas, "Where do i move my sensors? emergence of a topological representation of sensors poses from the sensorimotor flow," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–1, 2019.
- [16] J.-M. Godon, S. Argentieri, and B. Gas, "A formal account of structuring motor actions with sensory prediction for a naive agent," *Frontiers in Robotics and AI*, vol. 7, p. 179, 2020. [Online]. Available: <https://www.frontiersin.org/article/10.3389/frobt.2020.561660>
- [17] D. Pierce and B. J. Kuipers, "Map learning with uninterpreted sensors and effectors," *Artificial Intelligence*, vol. 92, no. 1, pp. 169–227, 1997.
- [18] L. A. Olsson, C. L. Nehaniv, and D. Polani, "From unknown sensors and actuators to actions grounded in sensorimotor perceptions," *Connection Science*, vol. 18, no. 2, pp. 121–144, 2006.
- [19] N. Le Hir, O. Sigaud, and A. Laflaquière, "Identification of invariant sensorimotor structures as a prerequisite for the discovery of objects," *Frontiers in Robotics and AI*, vol. 5, p. 70, 2018.
- [20] A. Laflaquière, S. Argentieri, O. Breysse, S. Genet, and B. Gas, "A non-linear approach to space dimension perception by a naive agent," *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Oct 2012, pp. 3253–3259.
- [21] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/036402139090002E>
- [22] J. McClelland, *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*, 2nd ed., Dec. 2015. [Online]. Available: <https://web.stanford.edu/group/pdplab/pdphandbook/handbook.pdf>
- [23] J.-M. Godon, "A structuralist formal account for sensorimotor contingencies in perception," Ph.D. dissertation, Sorbonne Université, 2022.
- [24] E. W. Dijkstra et al., "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.
- [25] J. B. Kruskal, "Nonmetric multidimensional scaling: A numerical method," *Psychometrika*, vol. 29, no. 2, pp. 115–129, 1964.
- [26] —, "Multidimensional scaling by optimizing goodness of fit to nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.
- [27] I. Borg and P. J. Groenen, *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.
- [28] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [29] D. Endres and J. Schindelin, "A new metric for probability distributions," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1858–1860, 2003.
- [30] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, March 1951.
- [31] D. H. Ballard and R. Zhang, "The hierarchical evolution in human vision modeling," *Topics in Cognitive Science*, vol. 13, no. 2, pp. 309–328, 2021.