



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-03537409v2

HAL Id: hal-03537409

<https://hal.science/hal-03537409v2>

Preprint submitted on 21 Jun 2022 (v2), 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 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 action 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 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 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 on 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,

etc.) to experiences that are themselves close: 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 attempted internally establishing 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 on generic topological intuitions, they all end up exploiting a discrete setup for which most of the topological structures are useless. Indeed, most of modern robotic setups rely on discrete time computations for which one can define other tools like distances based on similarities or correlations between elements in the agent sensorimotor flow. Then, should one 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 comparison at a “symbolic” level denoted by a strict equality operator between e.g. sensory values. While the previous cited contributions certainly proved 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

* 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.

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

105 Importantly, most of the previously cited contributions also
 106 claim to deal with uninterpreted sensory signals. But in these
 107 works, the form assumed by the signal (and the expected
 108 transformations thereof) is usually known and leveraged by
 109 the agent; what it ignores instead is *how* these signals re-
 110 late to the sensorimotor interaction. Then, on the basis of
 111 a priori distances, metrics, similarities, are built maps or
 112 representations of the agent’s sensorimotor interaction with
 113 a generally unknown environment [17]. In this paper, how-
 114 ever, no “natural” metrics nor algebraic operations on the
 115 symbols perceived by the agent are used –pretty much in
 116 line with [18] where a less formal approach is proposed–,
 117 contrary to our understanding of the usual numeric values.
 118 This is generally made manifest with the choice to assume
 119 that states are coded as numeric values (or tuples thereof),
 120 and of special influence with that of whether to use natural
 121 (possibly topological, as previously outlined) structures of \mathbb{R}^N .
 122 Thus, most developments which try to achieve robustness and
 123 scalability do so via extrapolation and clustering [12], [19],
 124 [14]. Nevertheless, as already argued, these techniques require
 125 referring to preexisting external metrics, which constitutes
 126 assumption about a priori knowledge we would like to avoid.
 127 In this paper, we then propose to examine how some notion
 128 of closeness –that we could also relate to some internal
 129 notion of *subjective* continuity– in sensorimotor experiences
 130 can emerge from uninterpreted sensory information for a naive
 131 agent operating in discrete time. To that end, some formal
 132 considerations are first introduced in §II. After evaluating a
 133 purely topological approach, a metric approach is proposed
 134 instead and the probability of transition between sensory
 135 symbols is used to define some appropriate notion of sensory
 136 distance. On this basis some simple simulations are introduced
 137 in Section III to illustrate how an agent could leverage some
 138 structure by simply judging if its sensory observations are
 139 close or not. This is illustrated for visual perception through
 140 the building by a naive agent of the grayscale or some RGB
 141 color model. Then, the role of the agent’s action in this
 142 framework is questioned in §IV. More precisely, the spatial
 143 and temporal dynamics of the agent’s exploration is shown
 144 critical to obtain a meaningful and useful structure of its own
 145 sensory symbols. Next, some experiments initially proposed
 146 in [16] are reproduced in Section V to illustrate how the
 147 proposed framework could allow an agent to actually build
 148 some sensory prediction functions even in the presence of

sensory noise. Finally, a conclusion ends the paper. 149

II. TOWARDS A TOPOLOGY OF SENSORY VALUES 150

This first section aims at defining a topology of sensory val- 151
 ues, built on the basis of the agent’s sensorimotor experience. 152
 After a short subsection devoted to the required definitions 153
 and notations, a time variable is added to the formalism in 154
 the second subsection, so as to account for explicit time 155
 dependency of the agent’s experience, allowing us to introduce 156
 a first time-inherited topology. While being possibly sufficient, 157
 arguments for the introduction of an explicit metric are then 158
 discussed. The third subsection thus proposes the definition of 159
 an internal probabilistic metric and highlights the benefits and 160
 limits of the proposed approach. Section 3 then exploits these 161
 elements in a simple experimental framework to illustrate these 162
 elements and demonstrate their actual exploitation. 163

A. A short reminder on notations 164

Let us consider in the following an agent endowed with 165
 motor and sensory capabilities. Its internal sensorimotor con- 166
 figuration is classically noted as (\mathbf{m}, \mathbf{s}) , where $\mathbf{m} \in \mathcal{M}$ (resp. 167
 $\mathbf{s} \in \mathcal{S}$) represents the agent’s internal motor (resp. sensory) 168
 configuration as an element of its corresponding motor \mathcal{M} 169
 (resp. sensory \mathcal{S}) set. As shown in [16], the agent’s motor 170
 description can be enriched from $\mathbf{m} \in \mathcal{M}$ to $\mathbf{b} \in \mathcal{B}$, where 171
 $\mathbf{b} = (\mathbf{m}, \boldsymbol{\tau})$ depicts the absolute agent’s motor configuration. 172
 \mathbf{b} is made of the agent’s internal (and thus known to it) 173
 motor configuration \mathbf{m} and of its absolute external (and thus 174
 unknown from it) pose $\boldsymbol{\tau}$ in its ambient space. Importantly, 175
 as discussed in [16], switching the motor description from 176
 \mathbf{m} to \mathbf{b} allows to keep a functional relation between motor 177
 and sensory data, even in the case where the agent can freely 178
 move in its environment. But while the agent has no direct 179
 access to \mathbf{b} , it can apply some *motor actions* a on $\mathbf{b} = (\mathbf{m}, \boldsymbol{\tau})$ 180
 to go to configurations $\mathbf{b}' = (\mathbf{m}', \boldsymbol{\tau}') = ab$: the agent 181
 knows instead how to *move* in \mathcal{B} . This capability will be 182
 exploited later to apply the following developments to get an 183
 internal assessment of sensory regularity, see §V. Next, the 184
 environment state is characterized as a function $\epsilon : \mathcal{X} \rightarrow \mathcal{P}$, 185
 i.e. as a state $\epsilon \in \mathcal{E}$ linking the ambient geometrical space \mathcal{X} 186
 in which sensorimotor experiences occur (classically endowed 187
 with some rigid transformations group $\mathcal{G}(\mathcal{X})$) to the set of 188
 the physical properties \mathcal{P} observable by the agent, where \mathcal{E} 189
 denotes the set of environmental states. Then, $\epsilon(\mathbf{x})$ represents 190
 the observable physical properties at point $\mathbf{x} \in \mathcal{X}$. On the 191
 basis on the previous definitions, one can now define the 192
 sensorimotor map ψ as the function $\psi : \mathcal{B} \times \mathcal{E} \rightarrow \mathcal{S}$, such 193
 that $\mathbf{s} = \psi(\mathbf{b}, \epsilon)$. One can notice here that the sensorimotor 194
 law does not explicitly depend on time, as is the case of most 195
 other contributions in the fields [10], [20], [14]. We will now 196
 enrich this formalization with an explicit time dependency. 197
 It will then constitute our gateway towards continuity in the 198
 sensory experience of the agent, much as in J. Elman’s famous 199
 1990 paper [21] where words are not considered as preexisting 200
 categories but more as emergent features in the latent structure 201
 of sentences along time [22], similarly to the topology of 202
 sensory symbols can be considered as the latent structure 203
 between them through the sensorimotor experience along time. 204

205 *B. All is well in continuous land*

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

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

212 With this notation, one can express an instantaneous snapshot
213 of the environmental state as the partial function

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

214 Therefore any temporal succession of environment states can
215 be described as a trajectory

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

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

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

218 and

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

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

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

succession continuous. So how can we solve this issue? One
251 proposes to turn to the setting of metric geometry, which
252 although less general is more suited, in the next subsection. 253

C. Introduction of a statistical sensory metric 254

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

Let us now define the likelihood $P_{s'|s}$ over all experiences
281 that the sensory value s' immediately follows s in the senso-
282 rimotor flow of the agent along 283

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

Importantly, from the previous time homogeneity assumption, 284
 $P_{s'|s}$ does not depend on the current time t it is computed. 285
From there and following the intuition that “closeness” of sen- 286
sory values \mathbf{s} and \mathbf{s}' should increase whenever the probability 287
of the transition $\mathbf{s} \rightarrow \mathbf{s}'$ does, we propose to define a simple 288
metric prototype via 289

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

where f should verify the two conditions: 290

- 1) $f : [0; 1] \rightarrow \mathbb{R}_+$: f only needs to map probabilities in 291
[0; 1] to nonnegative values, i.e. dissimilarity values; 292
- 2) f is non-increasing: probable transitions (i.e. $P_{s'|s}$ close 293
to 1) should result in low dissimilarities. 294

These conditions do not make δ_f a metric since it only verifies 295
the non-negativity property. We therefore extend it via minimal 296
paths considerations, i.e. by defining a distance d_f . Let $\mathcal{R}^{\mathbf{s}, \mathbf{s}'}$ 297
be the set of all paths from \mathbf{s} to \mathbf{s}' , with 298

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

299 We can then define d_f along

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

300 This in turn enforces the properties of *triangular inequality*
 301 and *reflexivity*. In the case where \mathcal{S} is finite, this reduces to
 302 the familiar computational form of finding minimal paths on
 303 a finite graph with nonnegative weights (corresponding to the
 304 $\delta_f(\mathbf{s}, \mathbf{s}')$ edge from \mathbf{s} to \mathbf{s}'). One should also note that this does
 305 *not* guarantee *symmetry* at its core because $P_{\mathbf{s}'|\mathbf{s}}$ may differ
 306 from $P_{\mathbf{s}|\mathbf{s}'}$. Then the δ_f weights naturally define a *directed*
 307 graph (*digraph*), which do not impair the search for minimal
 308 paths but do however lead to a non symmetric d_f function.
 309 While there exist several ways to obtain a closely related
 310 undirected graph from any given digraph, we hypothesize
 311 instead that symmetry should occur as a contingency of
 312 the sensorimotor exploration in most real world examples.
 313 Therefore, we do not enforce such corrections for now and
 314 will instead assess this hypothesis in the resulting graph.

315 III. BUILDING THE SENSORY TOPOLOGY FROM STATISTICS

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

327 A. Experimental setup and sensory distance estimation

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

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

344 where $s_{ij}[n]$ represents the (i, j) sensel value at time n , row
 345 i and column j of the agent camera sensor. Introducing $g(\cdot)$
 346 in (10) allows to explain formally how a physical state of the
 347 environment (which can be envisaged here as the pixel values
 348 of the video) is turned into the internal sensory state of the
 349 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 sensations as written in Equation 10. Then, we define
 a $S \times S$ matrix $M = (m_{kl})_{k,l}$ counting all the transitions of
 sensels 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
 representing 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), one 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 on the Δ 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

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

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

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

396 B. Results for a grayscale perception

397 The $W \times H = 856 \times 480$ video used to conduct the
 398 experiments comes from a slightly stabilized camera filming
 399 an evening walk in Midtown New York City in the rain². It
 400 consists in a natural city scene filmed in real time from a first
 401 person point of view. A grayscale (cropped) preview of the
 402 video is shown in Figure 1a. To begin with, one will consider
 403 a function g , mapping the (R_{ij}, G_{ij}, B_{ij}) color coding of the
 404 video pixels v_{ij} to the sensel values $s_{ij} \in \llbracket 0; 255 \rrbracket$ of the
 405 agent, such that

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

406 where h is a function that can be tuned to artificially modify
 407 the agent's perception. Note that g acts here like an excitation
 408 function, and is thus supposed identical for all sensels. Two
 409 cases for h are discussed in the following: either $h() = \text{id}()$ in
 410 §III-B1, corresponding to the case where the agent grayscale
 411 perception exactly matches the grayscale version of the video,
 412 or $h() = \text{sawtooth}()$ for which the perception is altered on
 413 purpose to exhibit the folding of the agent internal represen-
 414 tation between black and white pixel values in the video, as
 415 detailed in §III-B2.

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

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

beginning of the experiment only (see Figure 1e for $n = 10^4$)
 and mainly concerns *distant* symbols with a very low transition
 probability, i.e. in the two corners of the Figure 1b.

b) *Computation of the distance matrix:* On the basis
 on the previous probability of transitions between symbols,
 one can compute the metric prototype in the form of the
 $S \times S$ matrix Δ whose elements are given by Eq. (14).
 Then, Dijkstra's algorithm [24] is performed on Δ 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 Δ)
 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. One 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, one recognizes the same diagonal
 pattern, which now corresponds to low distances. One 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, one 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 $f = \text{id}()$). One
 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 neighbor of each symbol in the
 internal metric; these neighbors are then linked together in
 the projection by an arrow drawn in the figure. Browsing
 the manifold by following these arrows allows to go from

²https://youtu.be/eZe4Q_58UTU

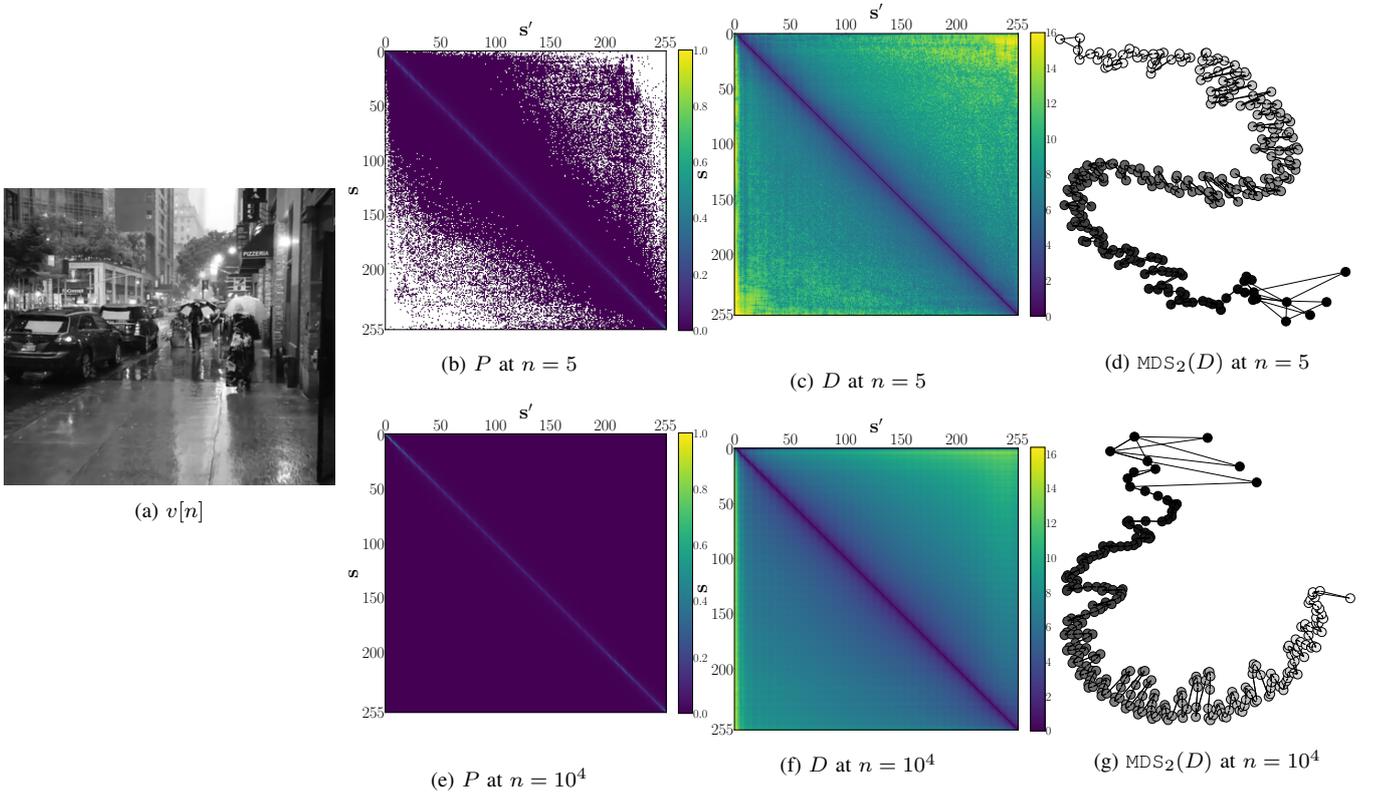


Figure 1: Building of the internal organization of sensory values. (a) Grayscale version of one frame of the video used in the experiment. (b)(e) Estimated probability matrix P at $n = 5$ and $n = 10^4$, i.e. at the very beginning of the experiment. (c)(f) Estimated distance matrix D at $n = 5$ and $n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

503 white (coded as the number 255) to black (coded as a 0)
 504 almost without any discontinuity in the symbol order at the
 505 end of the experiment. Interestingly, one can see that the
 506 projection obtained at the early stage of the experiment already
 507 exhibits a one dimensional manifold, with a thicker and less
 508 organized ordering of symbols though. Again, the contribution
 509 of all sensels to the same statistic certainly explains this nice
 510 quick convergence of the representation. Thus, from the final
 511 graph, one can conclude that the agent has been able, starting
 512 only from the probability of transition between uninterpreted
 513 sensory symbols, to discover the gray level scale. Such a
 514 capability will be further exploited for different applications,
 515 like sensory prediction, see §V.

516 2) 2nd case: $h()$ is a sawtooth function: We will now
 517 consider a case where the agent's sensory output does not
 518 exactly match the original grayscale world as per Eq. (16),
 519 where $h() = \text{sawtooth}()$ is defined along

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

520 for $x \in [0; 255]$ only. With such a change, a single internal
 521 sensory symbol (e.g. 54) will now correspond to two possible
 522 world grayscale values (27 and 155). Intuitively, such a change
 523 is expected to *create* continuity that does not exist initially
 524 between symbols through a stronger proximity between values
 525 representing dark and light shades. The previous process
 526 is then repeated and the resulting 2D MDS embedding is

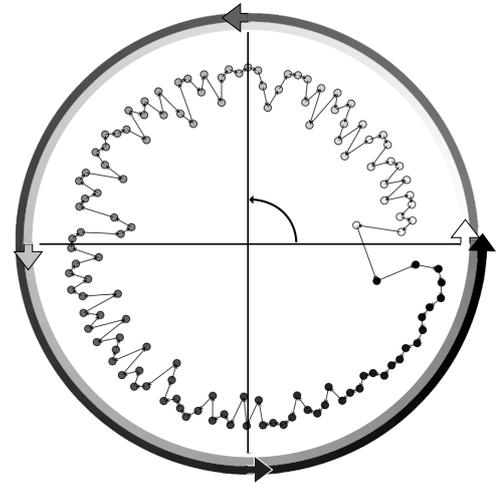


Figure 2: 2D MDS projection of the sensory symbols when a sawtooth function links together world gray values to sensory symbols. Each symbol is represented as a circle whose color represents the *internal coding*. The corresponding symbols *in the outside world* are represented as a looping arrow around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

depicted in Figure 2: as expected, one identifies a looping 527
 monodimensional manifold. In the figure, each sensory symbol 528
 is depicted as a circle whose color represents its *internal* 529

530 coding (i.e. a numerical value from 0 to 254 with a step of 2),
 531 represented as grayscale values for convenience. This color no
 532 longer matches the grayscale values of the world it represents
 533 because of the introduction of the sawtooth function. But the
 534 continuity initially captured in the previous experiment leads to
 535 a looping representation where the two opposite symbols 0 and
 536 254 are now close to each other in the internal representation
 537 as they both correspond to close grayscale values in the
 538 environment. Such a conclusion might be obvious in this
 539 specific case, but it highlights that the *internal, subjective*
 540 representation of the sensory symbols' topology might actually
 541 highly differ from our initial intuition as it depends on the
 542 way the agent's sensors encode sensory information. The same
 543 remark could apply to faulty sensors, which output symbols
 544 could be modified or rearranged because of some failure in
 545 the information acquisition process; the proposed approach
 546 could then allow the agent to (re)build an adequate internal
 547 representation, though still intrinsically limited by its own
 548 (possibly limited) sensory capabilities.

549 C. Results for color perception

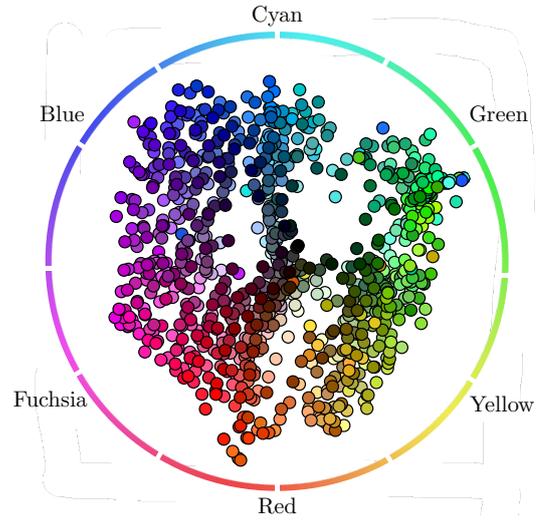
550 To further illustrate the approach, one will now endow the
 551 agent with some color perception capabilities. Then, in this
 552 subsection, the initial color tuples $(R_{ij}, G_{ij}, B_{ij}) \in \llbracket 0; 255 \rrbracket^3$
 553 coding the video pixels values v_{ij} are now mapped to the
 554 $S = \alpha^3$ agent's sensels values $s_{ij} \in \llbracket 0; \alpha^3 - 1 \rrbracket$ along

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

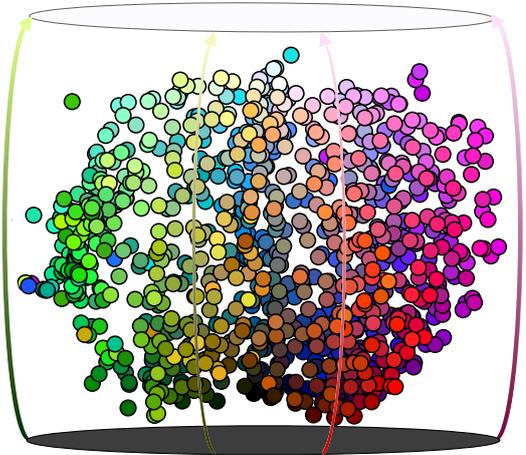
555 with $Q_\alpha(\cdot)$ a quantification function defined by

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

556 with $X \in \llbracket 0; 255 \rrbracket$ and $Q_\alpha(X) \in \llbracket 0; \alpha - 1 \rrbracket$. Note that while
 557 the symbol ordering was quite obvious for grayscale values
 558 from an external point of view (e.g. the natural order from 0 to
 559 255) for the various matrices M , P , Δ , and D , this no longer
 560 holds for these color sensory output symbols. Nevertheless, the
 561 order in which they appear as line or column indices in these
 562 matrices is not relevant since the only relevant information of
 563 closeness between them is entirely independent on how these
 564 symbols are actually ordered. In all the following, $\alpha = 10$
 565 is selected, so that the agent's sensory space is made of
 566 $S = \alpha^3 = 1000$ uninterpreted (numerical) symbols. On this
 567 basis, all the previous steps are successively applied. The
 568 resulting D matrix can then be visualized through a low
 569 dimensional embedding technique like ISOMAP [28]. The
 570 result of this projection performed in 3D is shown in Figure 3.
 571 The obtained representation is pretty much in line with some
 572 classical representations of RGB color models, like the HSL or
 573 HSV coding of color. Indeed, the 3D points cloud first appears
 574 to capture some color order very similar to the classical hue
 575 color wheel where pure colors are represented through an
 576 angular position on a circle, as depicted in Figure 3a. But
 577 the 3D projection also exhibits a third axis linking very dark
 578 to very light shades for each color of the hue wheel, similar
 579 to the lightness axis in the HSV color coding, see Figure 3b.
 580 In order to assess in a more quantitative way the similarity of



(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”: one can notice that all the sensory symbols are arranged by colors, 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 is supporting the variation of lightness. The obtained projection can thus be understood as analog to the HSL cylindre or biconic representation of the RGB color model.

this low dimensional projection with different color models, it is proposed to compute a Frobenius distances \mathcal{F} along

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

where D_m is the $S \times S$ distance matrix between all the sensory symbols observed during the experiment for the color model $m \in \{\text{RAND}, \text{RGB}, \text{HSL}, \text{HSV}\}$, and $|\cdot|$ the standardization operator. The resulting distances are reported in Table I, where the HSV color model better fits the obtained projection, as initially qualitatively intuited. 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

592 discontinuity in the low dimensional embedding, as a way to
 593 represent the information actually captured by the agent in D ,
 594 which is in the end the only data it exploits in the following.
 595 Then, with such a representation, the agent is now able to
 596 assess if the sensory symbol associated to the rose color is
 597 *closer* to the one associated to the red color than it is to the
 598 green one thanks to its internal metric matrix D .

599 IV. LOCOMOTIVE MOTIVES: A CASE FOR FITTING 600 EXPLORATION AND EXPLOITATION DYNAMICS

601 The previous developments were largely devoted to the
 602 relationship between two internal observations: the transition
 603 probabilities and the resulting metric. We showed how this
 604 information allows the agent to build some notion of closeness
 605 between sensory symbols –which could be understood as
 606 some subjective notion of sensory continuity– from certain
 607 successions of sensory experiences being more likely, or
 608 *typical*, than others. But such considerations clearly rely on the
 609 idea that typical environment states also display certain typical
 610 patterns themselves. From an external point of view, one would
 611 certainly declare that “environment states are (mostly) contin-
 612 uous”, both in time and space. This underlying assumption
 613 has not been dealt with so far, especially since the agent
 614 was passively observing sensory symbols changing along time
 615 in the previous experiments, and not actively exploring its
 616 environment. This section thus aims to study which external
 617 structure in the states of the environment could explain the
 618 relationships between the agent’s motor actions associated to
 619 a sensory experience and the observed regularity, effectively
 620 giving action a defining role in this internal assessment. To
 621 that end, additional formal considerations are introduced in
 622 a first subsection. On this basis, some new experiments are
 623 proposed to highlight the importance of movement in building
 624 this subjective continuity.

625 A. Fitting spatial and sensory dynamics in the exploration

626 1) *Spatial and temporal coherence*: The results obtained in
 627 Section III were based on a purely passive observation of a
 628 changing “natural” visual scene –where the word *natural* here
 629 refers to our own usual and intuitive sensorimotor experience–
 630 allowing the agent to build a representation of its own sensory
 631 symbols organization. But this representation should highly
 632 depend on the environment states and the successive config-
 633 urations with which the agent samples it along time. More
 634 precisely, this implies that the environment’s state should
 635 exhibit some typical patterns, both in space and time, in line
 636 with the manner in which the agent conducts its interaction, to
 637 make apparent the notion of certain sensory symbols transition
 638 being “more typical” than others. Thus, one first condition to
 639 fulfill is *spatial*, mandating that e.g. immediately next to a red

Color model	RAND	RGB	HSL	HSV
Distance \mathcal{F}	1320	947	890.4	857.2

Table I: Distances \mathcal{F} between the low dimensional projection and the corresponding color model, with RAND representing a random organization of the observed color symbols.

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 to another when Δt remains sufficiently small.
 One 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 appears
 two kilometers, 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
 $\mathcal{B}'_{\mathcal{T}} \subset \mathcal{B}_{\mathcal{T}}$,

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

for which $P_{s'|s}^{\mathcal{B}'_{\mathcal{T}}} = P_{s'|s}$. Such a (slight) generalization allows
 to highlight how a specific set of motor trajectories actually
 condition the sensory transitions available in the agent’s sen-
 sorimotor flow. More precisely, we introduced in [16] the
sensor receptive field as the specific region of space which
 environment state suffices to fully determine the agent 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, the corresponding local
 environment states $\gamma_{\epsilon|F(\gamma_{\mathbf{b}}(t+1))}(t+1)$ and $\gamma_{\epsilon|F(\gamma_{\mathbf{b}}(t))}(t)$
 are 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 $\gamma_{\mathbf{b}} \in$
 $\mathcal{B}'_{\mathcal{T}}$, do 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 $\gamma_s(t) = \gamma_{\gamma_{\mathbf{b}}(t), \epsilon|F(\gamma_{\mathbf{b}}(t))}(t)$
 are 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

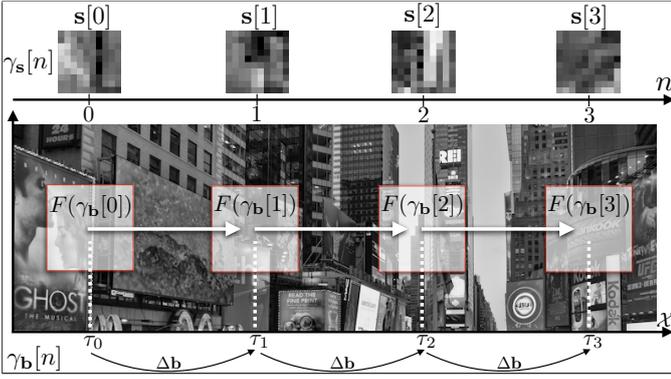


Figure 4: Experimental setup to assess the effect of the agent’s movement in the internal sensory symbol topology. A camera, whose field of view –or receptive field– is drawn as a square with red borders, faces a grayscale image and moves from one position τ to another thanks to an action of amplitude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

not verified– does not depend on previous sensory output \mathbf{s} anymore; instead, it simply replicates the *unconditional* probability that the agent experiences the particular sensory value \mathbf{s}' . To the agent, this means that the knowledge of which sensation \mathbf{s} it experiences at timestep t does not give it *any* information on which sensation \mathbf{s}' it is poised to experience at $t + 1$. Importantly, this shows that suitable choices of motor explorations are required for building a valid sensory metric, as well as giving an internal observation to assess whether this condition fails through Equation (25). The influence of this motor exploration is experimentally studied in the next subsection to illustrate these developments.

B. An experimental assessment of the influence of the movement amplitude

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

1) *Experimental setup*: let us consider in all the following a very simple agent, whose body is made of a planar, rectangular, camera sat atop one actuator allowing the agent to only move in one direction, see Figure 4. The pixels of the camera are sensitive to the luminance of the ambient stimulus, which is a fixed grayscale image placed in front of the moving camera. In such a case, the ambient space \mathcal{X} is then the plane \mathbb{R}^2 , and the state of the environment is a function ϵ mapping a position (x, y) in the plane to luminance values $\epsilon(x, y) \in \llbracket 0; 255 \rrbracket$ as encoded in the grayscale image. Those values are then converted into a sensory vector $\mathbf{s} \in \llbracket 0; 255 \rrbracket^{W \times H}$ directly capturing the corresponding grayscale value in the environment (the function $h(\cdot)$ in Equation (16) is thus the identity function). In the forthcoming simulations, $W = H = 100$. As already outlined in §II-A, we consider the agent is able to move in its environment by applying a single *action* a [16], i.e. by applying a function a to its current absolute configuration

$\mathbf{b} = (\mathbf{m}, \tau)$ to go to another configuration $\mathbf{b}' = (\mathbf{m}', \tau') = a\mathbf{b}$. In this section, we will mainly study the influence of the amplitude $\Delta\mathbf{b}$ of this action, which is supposed to produce a movement of the camera in only one direction and with the same amplitude, as illustrated in Figure 4. This is obviously a very particular and restrictive action, at least in comparison with the more generic motor actions 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.

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 τ_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 symbols topology. The experience is finally repeated for various amplitudes $\Delta\mathbf{b}$.

2) *Results and discussions*: 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 its 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 amplitude, 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, one faces here more or less the same conditions than in §III-B1 where the observation of the environment changes matches the changes in perception induced by 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

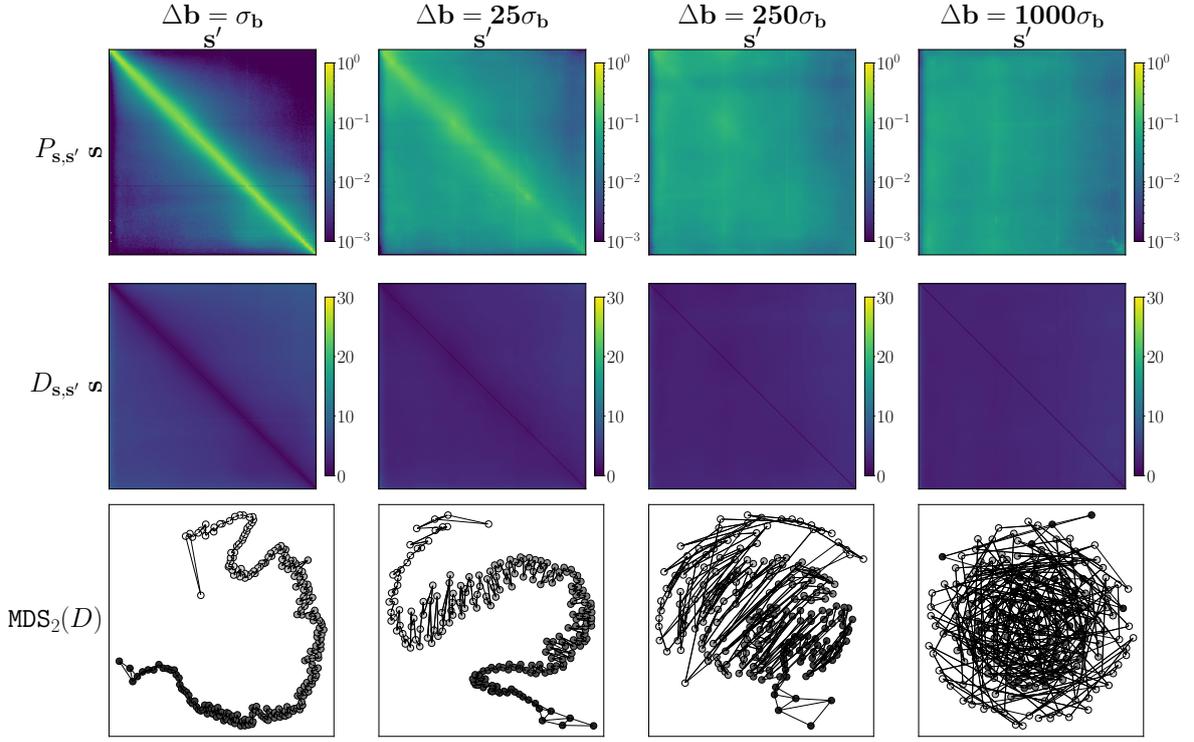


Figure 5: Evolution of the transition probability measure P (displayed with logarithmic norm), the distance measure D and the representation of this distance projected using 2-dimensional MDS for increasing movement amplitudes. Each column represents a motor trajectory for a fixed amplitude $\Delta \mathbf{b}$ described relatively to a 1 pixel shift of its sensor’s field of view. One can see that the diagonal pattern for P and D as well as the uni-dimensional grayscale manifold are deteriorating as the movement amplitude get 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.

when computing the distance matrix D from P (2nd row in Figure 5): for high amplitudes, mostly all symbols are now close to each other. Obviously, this results in very different $2D$ projections of the matrix D (3rd row). For lowest amplitudes, one still clearly sees a one-dimensional manifold, folding on itself when the action amplitude grows. But the dimensionality of the manifold is not sufficient to tell if the agent correctly captured or not the sensory symbols topology. Like we did in Figure 1g and Figure 1d, a k -NN algorithm is computed on D and displayed in Figure 5 to link each symbol to its closest neighbor, this link being represented as an arrow between two symbols in the projection. Looking at the smallest amplitude, the sensory symbols manifold can be browsed in the usual grayscale order by following the aforementioned arrows. On the contrary this proves impossible for larger amplitudes, where arrows link together e.g. symbols associated to clear and dark gray levels. It is then clear that the conditions written as Equations (22) and (21) are not verified anymore, with two successively sampled environment states associated to two distant positions in space, leading to the loss of perception by the agent of the spatial and temporal coherence in the environment.

Equation (25) states that, for a specific set of motor trajectories $\mathcal{B}'_{\mathcal{T}}$ making successive receptive fields falling apart from one another, the probability of transition between successive sensory symbols tends to an unconditional probability $P_{s'|s}^{\mathcal{B}'_{\mathcal{T}}} = P_{s'}$. Importantly, this phenomenon can be internally

assessed by the agent since both probability distributions are only based on sensory symbols observations; this then constitutes some *internal* way for the agent to rate the spatial and temporal coherence of its interaction with the environment. To that end, we propose to compare the two probability distributions $P_{s'|s=s_k}$ –the probability of every sensory value to succeed to a specific sensory value s_k – and $P_{s'}$, by using the Jensen-Shannon distance D_{JS} [29], a bounded metric based on the symmetrized version of the Kullback–Leibler divergence [30], and defined along

$$D_{JS}(P_{s'|s=s_k} \| P_{s'}) = \sqrt{\frac{\text{KL}(P_{s'|s=s_k} \| M) + \text{KL}(P_{s'} \| M)}{2}}, \quad \text{with } M = \frac{1}{2} (P_{s'|s=s_k} + P_{s'}), \quad (26)$$

with the KL divergence for two discrete probabilistic distributions A and B defined on the probability space \mathcal{W} as

$$\text{KL}(A \| B) = \sum_{x \in \mathcal{W}} A(x) \log_2 \left(\frac{A(x)}{B(x)} \right). \quad (27)$$

This results in a distance $D_{JS}(\cdot)$ between 0 and 1, computed for each sensory symbol s' , that is expected to converge towards 0 when both distributions are identical, i.e. when the motor trajectory of the agent leads to having s and s' independent. Again, D_{JS} is computed for 4 different amplitudes $\{\sigma_b, 5\sigma_b, 25\sigma_b, 125\sigma_b\}$ with corresponding graphs in

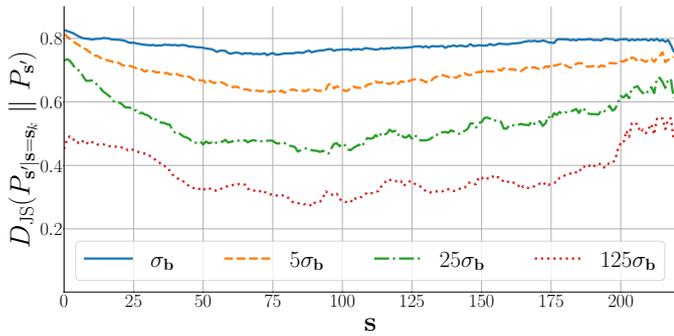


Figure 6: JS distance of every conditional probabilities relatively to the unconditional probability $P(s')$ for different movement amplitudes. Each point of the plot represents a JS distance for a single conditional probability to $P(s')$. As the amplitude increases the divergence of every symbol decreases,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

Figure 6. The results displayed in Figure 6 show that the JS distance for every probability distribution systematically decreases when the amplitude of agent's movement increases, i.e. the conditional distribution tends towards the unconditional one as described by Eq. (25). In the same vein, one can also conduct this comparison by computing the mutual information between, roughly speaking, the sensory symbols before and after the agent 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 of 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, this 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 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 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 a first subsection. Then, a 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 f^{-1} be; 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

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

One then has 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” τ_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. $d_f(\mathbf{s}, \mathbf{s}') < \tau_r$).

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 δ_{kl} . In such a case, still with a threshold τ_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.

910 B. Exploiting the sensory regularity for sensory prediction

911 We now propose to exploit the agent's capability to decide
 912 whether a sensory transition is regular in a sensory prediction
 913 task in the presence of noise inside the sensory data. To that
 914 end, the framing of the approach in [16] is first shortly intro-
 915 duced, followed by the proposed experimental setup, mirroring
 916 that of this previous contribution. Then, the sensory prediction
 917 framework from [16] is applied for different scenarios: (i) with
 918 no noise in the sensory data or with noise, but (ii) without or
 919 (iii) with rating the sensory regularity. A discussion comparing
 920 these scenarios is then proposed in a second paragraph.

921 1) *A short recall on the framing of the problem:* The con-
 922 tribution from [16] is all about the theoretical conditions for
 923 the determination of a sensory prediction function for a naive
 924 agent. More precisely, it is demonstrated how the algebraic
 925 structure found in this prediction is homeomorphic to that of
 926 an algebraic group of specific motor actions, the *conservative*
 927 actions. An action a is said conservative if all sensels of
 928 the agent exchange the places they sample when applying a :
 929 equivalently, conservative actions can then be thought of as
 930 permutation of sensels. Importantly, this result has since been
 931 extended to *quasiconservative* actions in [23], where partial
 932 sensory prediction maps are proposed to generalize the sensel
 933 permutations of strictly conservative actions for the case where
 934 some sensels have no identified permutations when applying
 935 an action a (e.g. for sensels in the border of a camera).

936 2) *Experimental setup:* the proposed simulation setup is
 937 very close to the one already presented in §IV-B1. The agent
 938 is still made of a moving camera facing a fixed grayscale
 939 image, as shown in Figure 4. This time, the agent is endowed
 940 with a $W \times H = 10 \times 10$ sensor, and is now able to
 941 move in four orthogonal directions by applying 5 different
 942 (quasiconservative) actions a_{id} , a_f , a_b , a_r and a_l making the
 943 camera receptive field respectively stay still, move in the left,
 944 right, up or down directions in \mathcal{X} . It is clear that the sensory
 945 consequences of such actions can be illustrated as a slip of
 946 information in the image in the opposite direction of the agent
 947 movement: most of the sensels values *before* applying any of
 948 these actions can found a successor *after*. Then, predicting
 949 the sensory consequence of an action can be summed up by
 950 a permutation between sensels values, providing all sensels
 951 share the same excitation function, as already outlined in
 952 §III-A2. Importantly, the agent has no clue about the incidence
 953 of a given action nor about their possible relationship. All it
 954 can do is perform an action and observe its consequence in its
 955 sensory data [16]. The proposed experimentation then relies
 956 on the two following steps.

957 a) *Step 1: building of the sensory symbol topology.* The
 958 agent explores its environment by randomly selecting an action
 959 in $\mathcal{A} = \{a_{id}, a_f, a_b, a_r, a_l\}$ with identical amplitudes $\Delta \mathbf{b} = \sigma_{\mathbf{b}}$
 960 (apart from a_{id}), and then infers the distance matrix D , in line
 961 with §IV-B where the number N_a of draws of actions is set
 962 to $N_a = 25$, and the number of repetition is selected to $N_r =$
 963 $2 \cdot 10^3$. As opposed to the previous case, some artificial noise
 964 n_{ij} is now added to the pixel value v_{ij} of the image to form

the agent's sensel values³ $s_{ij} = v_{ij} + n_{ij}$ 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^{th} sensel takes the value of
 the l^{th} sensel after applying action a_p . For this experiment,
 $N_a = 50 \cdot 10^3$ and $N_r = 1$. Initially, every element $\pi_{kl}^{(p)}$ of
 the permutation matrices Π_{a_p} is initialized to 1, meaning that all
 permutations between the agent sensels are possible for action
 a_p . Then, each time this action is drawn from \mathcal{A} , the agent
 can discard in Π_{a_p} some permutations by observing that some
 sensels values do not switch with one others, 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 to the
 element at the position (k, l) in the permutation matrix Π_{a_p} .
 One 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. Benefiting from the previous
 developments, one instead proposes a revised update rule of
 the permutation matrix along

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

where τ_r is a manually chosen threshold applying on the
 built matrix distance D . In the following, τ_r is tuned so
 as to correspond to the smallest threshold that allows for
 permutations matrices to converge. It is clear that this 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 veine, this second step in this experiment requires
 a second exploration of the *same* environment than during the
 first step. This is for sure 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 the
 building the sensory prediction functions. Importantly, this
 again highlights the importance of the threshold τ_r which
 should then also be selected by an appropriate combination
 of these two steps.

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

1012 *c) Evaluation: convergence of the permutation matrices.*
 1013 To evaluate the influence of the added noise on the conver-
 1014 gence of permutation matrices Π_{a_p} , one proposes an (external)
 1015 criterion $C(\Pi_{a_p}) = C_H(\Pi_{a_p}) \times C_D(\Pi_{a_p})$ adapted from [16]
 1016 to account for the added noise to the data and defined along

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

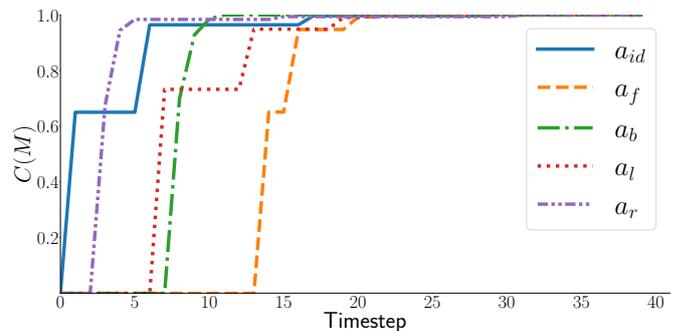
$$C_H(\Pi_{a_p}) = 1 - \frac{1}{N_s \log_2(N_s)} \sum_{i=1}^{N_s} H_i,$$

1017 with

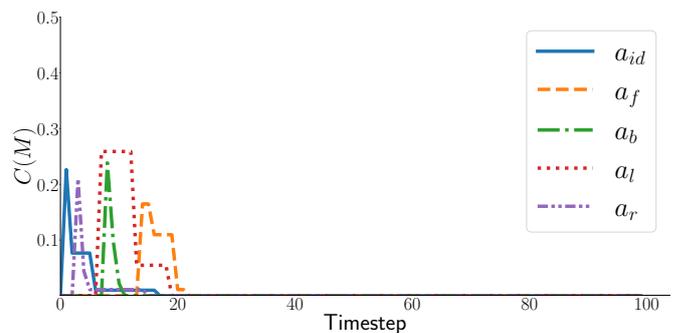
$$\begin{cases} H_i &= -\sum_{l=1}^{N_s} \frac{\pi_{kl}^{(p)}}{\mu_k} \log_2 \left(\frac{\pi_{kl}^{(p)}}{\mu_k} \right), \\ \mu_k &= \max \left(1, \sum_{l=1}^{N_s} \pi_{kl}^{(p)} \right), \end{cases} \quad (35)$$

1018 where $\bar{\pi}_{kl}^{(p)}$ represents the (binary) coefficients of the ideal
 1019 matrix $\bar{\Pi}_{a_p}$ associated to the action a_p . Basically, C_H can be
 1020 understood as an average measure of certainty in the discovery
 1021 of the permutations, weighted by the percentage C_D of the
 1022 correctly identified permutations w.r.t. the ground truth to
 1023 account for the noise possibly discarding some of them. In the
 1024 end, criterion C lies between 0 –i.e. the matrix is full of 1’s
 1025 (initialization) or 0’s (all permutations have been discarded)–
 1026 and 1 –i.e. the permutation has been correctly discovered–.

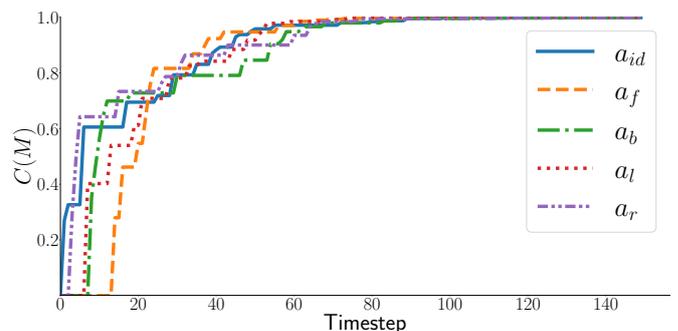
1027 *3) Results:* As outlined in the introduction of Section V,
 1028 three different scenarios are evaluated. To begin with, one
 1029 first considers the case where there is no noise in the agent’s
 1030 perception by setting $\sigma_n = 0$. Then, using the update rule (32)
 1031 should allow the agent to correctly build all of its permutation
 1032 matrices, exactly as in [16]. As expected, Figure 7a shows
 1033 that criterion C converges towards its maximal value 1 for
 1034 all actions in \mathcal{A} . C plots also exhibit sparse jumps at random
 1035 times, corresponding to the steps where the action was actually
 1036 drawn in \mathcal{A} during the experiment. More importantly, one can
 1037 see in Figure 7a that only a few realizations of each action a_p
 1038 (about 4 to 6 here) is required for $C(\Pi_{a_p})$ to almost reach 1,
 1039 showing how easy it is for the agent to discover the existence
 1040 of such permutations in its own perception. In the second
 1041 scenario, a noise of amplitude $\sigma_n = 2$ is now added to the
 1042 sensation. Obviously the strict comparison of sensel values
 1043 in (32) in the presence of such noise (however small) entirely
 1044 breaks the approach, as shown in Figure 7b. As expected, the
 1045 criterion C now converges to 0: each Π_{a_p} matrices converges
 1046 to null matrices as all possible permutations of values have
 1047 been (including erroneously) discarded in the process. Finally,
 1048 the new update rule (33) is now used to judge of the closeness
 1049 of sensel values on the basis on the built distance D , resulting
 1050 in the evolution of the criterion C represented in Figure 7c.
 1051 For this scenario, $\sigma_b = 1$ and $\tau_r = 1.63$. In the presence of
 1052 noise, the ability for the agent to assess if a sensation is now
 1053 close to others allows it to correctly discover the existence of
 1054 permutations in its perception. But clearly, this task is not as
 1055 easy as in the first scenario: the number of required actions for
 1056 correctly evaluating their corresponding permutation matrices
 1057 is significantly higher. This is apparent in Figure 7c, not only in
 1058 the slower convergence time of the criterion C , but also in the



(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 realization 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 τ_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).

smaller jumps of values in C . Indeed, each generation of action 1059 brings less information in the prediction process because of the 1060 noise included in the agent sensation. But still, the important 1061 structures anchoring the sensorimotor interaction the agent has 1062 with its environment are still available, allowing it e.g. to build 1063 an image of its body [14] or of its peripersonal space [15], at 1064 least at the cost of a longer interaction in time. 1065

VI. CONCLUSION 1066

In this paper, and after purely topological considerations, 1067 a metric-based approach is proposed to formalize the ability 1068 for a naive agent to build some subjective sense of sensory 1069

continuity. An experimental framework is then proposed, illustrated and assessed in the context of visual perception for the discovering of gray or color scales. Then the importance of the dynamic of the agent exploration relatively to that of the environment is studied, highlighting an important spatiotemporal coherence principle of this exploration. Finally, a sensory closeness notion being now available to the agent, a sensory 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 one 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 which are not directly correlated to its own action. 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. Dennhardt, 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. 1126
- [9] A. Censi, "Bootstrapping vehicles: A formal approach to unsupervised sensorimotor learning based on invariance," Ph.D. dissertation, California Institute of Technology, June 2012. 1127–1129
- [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. 1130–1132
- [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. 1133–1134
- [12] A. Laflaquiere, "Grounding the experience of a visual field through sensorimotor contingencies," *Neurocomputing*, vol. 268, no. C, 2017. 1135–1137
- [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> 1138–1139
- [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. 1140–1143
- [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. 1144–1146
- [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> 1148–1151
- [17] D. Pierce and B. J. Kuipers, "Map learning with uninterpreted sensors and effectors," *Artificial Intelligence*, vol. 92, no. 1, pp. 169–227, 1997. 1152–1153
- [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. 1154–1155
- [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. 1156–1159
- [20] A. Laflaquière, S. Argentieri, O. Breyse, S. Genet, and B. Gas, "A non-linear approach to space dimension perception by a naive agent," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Oct 2012, pp. 3253–3259. 1160–1163
- [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> 1164–1166
- [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> 1167–1169
- [23] J.-M. Godon, "A structuralist formal account for sensorimotor contingencies in perception," Ph.D. dissertation, Sorbonne Université, 2022. 1170–1172
- [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. 1173–1174
- [25] J. B. Kruskal, "Nonmetric multidimensional scaling: A numerical method," *Psychometrika*, vol. 29, no. 2, pp. 115–129, 1964. 1175–1176
- [26] —, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964. 1177–1178
- [27] I. Borg and P. J. Groenen, *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005. 1179–1180
- [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. 1181–1183
- [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. 1184–1185
- [30] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, March 1951. 1186–1188
- [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. 1189–1191