Where do I move my sensors? Emergence of a topological representation of sensors poses from the sensorimotor flow

Valentin Marcel, Sylvain Argentieri and Bruno Gas

***Abstract*—This paper deals with the perception of mobile robotic systems within the framework of interactive perception, and inspired by the sensorimotor contingencies (SMC) theory. These approaches state that perception arises from active exploration of an environment. In the SMC theory, it is postulated that information about the structure of space could be recovered from a quasi-uninterpreted sensorimotor flow. In a recent article, the authors have provided a mathematical framework for the construction of a sensorimotor representation of the interaction between the sensors and the body of a naive agent, provided that the sensory inputs come from the agent's own body. An extension of these results, with stimulations coming from an unknown changing environment, is proposed in this paper. More precisely it is demonstrated that, through repeated explorations of its motor configurations, the perceived sensory invariants can be exploited to build a topologically accurate internal representation of the relative poses of the agent's sensors in the physical world. Precise theoretical considerations are provided as well as an experimental framework assessed in simulated but challenging environments.**

I. Introduction

Space perception is a central issue in mobile robotics. Indeed, many abilities heavily depend on it as they are deeply rooted on spatial knowledge: among others, one can cite trajectory planning [1], obstacle avoidance [2], auditory and visual sensing [3], etc. Most of these implementations consider that space is something that exists objectively *out there*, and try to exploit it to model mechanical systems, localize some objects of interest, reach and catch moving targets, etc. However, space has not to be a pre-established substrate *per se* to be able to perform the very same tasks. In the case of the sensorimotor contingencies theory (SMC) [4], [5], it is claimed that space is something that an agent may experience via the determination of sensorimotor invariants called *contingencies*. In other words, the discovery at first, and then the use of such contingencies, is enough to make an agent realize actions without the need of having an internal, local or global representation, analytic or not, of space. A. V. Terekhov and J. K. O'Regan [6] have shown it unambiguously by learning from the sensorimotor flow an internal function representing translations. Since then, this idea has been extended by Le Clec'H et al. [7] where an agent exploits compensable sensory changes to build an internal representation of two-dimensional rigid transformations. The underlying idea of these works is based on the notion of active *compensable sensory changes* proposed initially by Poincaré [8], [9]. Since then, substantial works have been dealing with how action can structure perception [10]–[14], with a focus on Poincaré's idea on the notion of space, and more recently with growing interest in robotics applications. In this vein, Bohg et al. [15] has introduced the notion of "interactive perception" as the set of approaches in robotics concerned with the implication of action in perception.

In this interactive perception framework, robotics approaches to SMC appear to be more about finding geometrical properties of space through the exploitation of the sensorimotor flow; these approaches are not, for now, about compensable transformations as in Poincaré's intuition, but are still related to space. Thus, Philipona's first formalization of the SMC approach [16] led to the demonstration that an agent can, without any a priori, infer the so-called "dimension of space". Further works like [17] have also shown that, beyond the dimension of space, it is possible to build a motor internal representation of the positions occupied by the agent's end-effector without external knowledge about its working space. Despite the use of a curvilinear component analysis (CCA) [18] and the definition of adapted Hausdorff distances in the agent motor space, this work lacks a proper mathematical formalization and there are no clear definitions of the properties or spaces that are actually captured by the agent. As a solution, the authors have demonstrated that topological properties of the space of sensory invariants on a sensitive body can also be well captured [19]. This work focused on the agent self-interaction with its own body, whose perception is more deeply analyzed by Laflaquière in [20] within the SMC theory framework. Working on the agent's body was initially envisaged as a way to put the environment dependency of the representation aside. Indeed, as formalized later in this paper, the fact that the environmental state can possibly evolve along exploration has already proven to be a major theoretical difficulty [17], [21]. This paper proposes to tackle the environment dependency by generalizing the formalism we initially proposed [19]. Differently from this previous contribution, it is envisaged that the agent's end-effector is not sampling its own body anymore in such a way that all the sensory inputs are dependent to the environmental state. Along the changes in the environmental states, it is shown that an agent can undergo, by successive iterations, a partitioning process of its motor set. Starting from an unstructured motor set, the agent will build incrementally a final motor partition which finally forms a good representation of the external working space initially unknown to the agent. As such, this *refinement* process is conducted along the agent's

V. Marcel, S. Argentieri and B. Gas are with Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, ISIR, F-75005 Paris, France

life, through an active exploration of sensory invariant in changing environments. The approach is similar to Shalizi and Crutchfield [22] with a spatial version by Capdepuy et al. [23], where statistics of the sensory stream are exploited to build a model that best predicts sensory inputs. However, in the current paper the representation is built without using prediction but by using naive motor exploration and statistics of the sensory invariants, which are obtained by hypothesizing that the agent is only able to detect when two sensations are equal; no other interpretation of the sensations is actually required. The idea of measuring spatial distances between sensors using statistics of uninterpreted sensory inputs has already been tackled by multiple papers [24], [25], with different metrics exploiting the sensory distributions in static environments. In this paper it is shown that, under some specific hypotheses on the environment dynamical statistics, the internal representation is topologically equivalent to the space of *sensitive* displacements of the sensors.

This paper is organized as follows. Section II is devoted to the mathematical formalization of the aforementioned refinement process of the agent motor set. Additionally to all the theoretical considerations, a simple example is used all along the section (and in the rest of the paper) to illustrate this process and its limits. Section III introduces some formal considerations on the topological and metrical structures in the sets obtained during the refinement process but also some criteria to evaluate how the topological structures are preserved in an internal representation. The theory is then exploited in Section IV to propose an experimental and probabilistic framework for building this internal representation. Simulations are conducted in Section V to carefully evaluate this framework. Finally, a conclusion ends the paper.

II. The refinement process: partitioning of the motor space through sensory invariants

A. Characterization of the motor space

1) Refinement of the motor space: Let's first consider a naive agent, be it virtual or robotic, that can interact with its environment by generating motor commands that act on the agent actuators states (i.e. joint angles, positions, etc.). Each state is called a *motor configuration* $\boldsymbol{m}$ and lies in the *motor configuration set* $\mathcal{M}$. The agent is also endowed with sensors placed on its body parts. These sensors inform the agent about the environment's physical state they are sensitive to, thus generating a *sensory input* $\boldsymbol{s} \in \mathcal{S}$, with $\mathcal{S}$ the *sensory set*. An environment's physical state is called an *environmental state* $\epsilon \in \mathscr{E}$ where $\mathscr{E}$ is the set of all possible environmental physical states, possibly infinite. It is also assumed that there exists a deterministic function Ψ, often referred to as the *sensorimotor law* [16], that links the motor configuration state $\boldsymbol{m}$, the environmental state $\boldsymbol{\epsilon}$ and the sensory input $\boldsymbol{s}$ as

$$\boldsymbol{s} = \Psi(\epsilon, \boldsymbol{m}) = \Psi_\epsilon(\boldsymbol{m}). \tag{1}$$

Note that time does not appear in this equation. Indeed, the environmental state ϵ is assumed to capture time changes in the environment, while the sensory input $\boldsymbol{s}$ is hypothesized to instantaneously reach its final value. Therefore, for a fixed environmental state, the sensory input is solely governed by the motor configuration $\boldsymbol{m}$. Because of the possible redundancies in the agent kinematics, the function $\Psi_{\boldsymbol{\epsilon}}(.)$ is possibly non-injective. It means that, at $\boldsymbol{\epsilon}$, two different motor states $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ can lead to the very same sensory input $\boldsymbol{s} = \Psi_{\boldsymbol{\epsilon}}(\boldsymbol{m}_1) = \Psi_{\boldsymbol{\epsilon}}(\boldsymbol{m}_2)$. As outlined in our previous work [19], one can then define an equivalence relation $=_{\Psi_{\boldsymbol{\epsilon}}}$, such that

$$\boldsymbol{m}_1 =_{\Psi_{\boldsymbol{\epsilon}}} \boldsymbol{m}_2 \Leftrightarrow \Psi_{\boldsymbol{\epsilon}}(\boldsymbol{m}_1) = \Psi_{\boldsymbol{\epsilon}}(\boldsymbol{m}_2). \tag{2}$$

Thus, one can regroup all the motor states leading to the same sensory state in their equivalence class $[\boldsymbol{m}]_{=_{\Psi_{\boldsymbol{\epsilon}}}} = [\boldsymbol{m}]_{\boldsymbol{\epsilon}} = \{\boldsymbol{r} \in \mathcal{M} | \boldsymbol{r} =_{\Psi_{\boldsymbol{\epsilon}}} \boldsymbol{m}\}$. It is well known that the set of all equivalence classes forms a *partition* [1] of the set on which the equivalence relation is defined. In other words, every element in $\mathcal{M}$ is included in one and only one equivalence class $[\boldsymbol{m}]_{\boldsymbol{\epsilon}}$. This partition is called the quotient set $\mathcal{M}/_{\boldsymbol{\epsilon}} = \{[\boldsymbol{m}]_{\boldsymbol{\epsilon}} | \boldsymbol{m} \in \mathcal{M}\}$. It forms a *refinement* of the trivial partition $\{\mathcal{M}\}$, i.e. the set $\mathcal{M}$ is split into equivalence classes of $\mathcal{M}/_{\boldsymbol{\epsilon}}$.

Consider now that the environmental state has switched from $\boldsymbol{\epsilon}$ to $\boldsymbol{\epsilon'}$. Then the agent has access to a new equivalence relation $=_{\Psi_{\boldsymbol{\epsilon'}}}$, that leads to a new motor set partition $\mathcal{M}/_{\boldsymbol{\epsilon'}}$. Because the environment has changed, the new equivalence relation and motor partition are possibly totally different from their counterpart in the previous environment. Therefore, one can define a new *multi-environment* equivalence relation $=_{\Psi_{(\boldsymbol{\epsilon},\boldsymbol{\epsilon'})}}$ such as

$$\boldsymbol{m}_1 =_{\Psi_{(\boldsymbol{\epsilon},\boldsymbol{\epsilon'})}} \boldsymbol{m}_2 \Leftrightarrow \begin{pmatrix} \boldsymbol{m}_1 =_{\Psi_{\boldsymbol{\epsilon}}} \boldsymbol{m}_2 \\ \text{and} \\ \boldsymbol{m}_1 =_{\Psi_{\boldsymbol{\epsilon'}}} \boldsymbol{m}_2 \end{pmatrix}. \tag{3}$$

According to its definition, this equivalence relation leads to equivalent classes $[\boldsymbol{m}]_{(\boldsymbol{\epsilon},\boldsymbol{\epsilon'})} = [\boldsymbol{m}]_{\boldsymbol{\epsilon}} \cap [\boldsymbol{m}]_{\boldsymbol{\epsilon'}}$ verifying $[\boldsymbol{m}]_{(\boldsymbol{\epsilon},\boldsymbol{\epsilon'})} \subseteq [\boldsymbol{m}]_{\boldsymbol{\epsilon}}$ and $[\boldsymbol{m}]_{(\boldsymbol{\epsilon},\boldsymbol{\epsilon'})} \subseteq [\boldsymbol{m}]_{\boldsymbol{\epsilon'}}$, i.e. the quotient set $\mathcal{M}/_{(\boldsymbol{\epsilon},\boldsymbol{\epsilon'})}$ is a refinement of both quotient sets $\mathcal{M}/_{\boldsymbol{\epsilon}}$ and $\mathcal{M}/_{\boldsymbol{\epsilon'}}$. Note that this multi-environment equivalence relation $=_{\Psi_{(\boldsymbol{\epsilon},\boldsymbol{\epsilon'})}}$ does not depend on the order of $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon'}$. Consequently, the tuple $(\boldsymbol{\epsilon},\boldsymbol{\epsilon'})$ can be written as a subset $E = \{\boldsymbol{\epsilon},\boldsymbol{\epsilon'}\}$ of $\mathscr{E}$. Based on the idea that intersecting partitions obtained on multiple environments gives a finer partition, one can then define the generic multi-environment equivalence relation $=_{\Psi_E}$, for any subset $E \subseteq \mathscr{E}$, as

$$\boldsymbol{m}_1 =_{\Psi_E} \boldsymbol{m}_2 \Leftrightarrow \ \Psi_{\boldsymbol{\epsilon}}(\boldsymbol{m}_1) = \Psi_{\boldsymbol{\epsilon}}(\boldsymbol{m}_2), \forall \boldsymbol{\epsilon} \in E. \tag{4}$$

Equality of sensory inputs must be valid for all environmental state in E, therefore, by indexing the obtained sensations by the corresponding environmental state, relation (4) can be rewritten as

$$\begin{aligned} &\boldsymbol{m}_1 =_{\Psi_E} \boldsymbol{m}_2 \\ \Leftrightarrow\ & \{(\epsilon, \Psi_{\boldsymbol{\epsilon}}(\boldsymbol{m}_1))\,; \boldsymbol{\epsilon} \in E\} = \{(\epsilon, \Psi_{\boldsymbol{\epsilon}}(\boldsymbol{m}_2)));\boldsymbol{\epsilon} \in E\} \\ \Leftrightarrow\ & \Psi_E(\boldsymbol{m}_1) = \Psi_E(\boldsymbol{m}_2), \end{aligned} \tag{5}$$

where the function

$$\Psi_E(\boldsymbol{m}) = \{(\epsilon, \Psi_{\boldsymbol{\epsilon}}(\boldsymbol{m})); \boldsymbol{\epsilon} \in E\} \in \mathcal{S}_E \tag{6}$$

[1] A partition of a set X is a set of non-empty, pairwise disjoints, subsets whose union forms the set X itself.

maps each motor configuration to its respective *indexed sensory set* acquired along the experience of all environmental states $\epsilon \in E$ and $\mathcal{S}_E$ is the set of all possible *indexed sensory sets* obtained from E. Experiencing new environmental states always give a more refined partition. Then, considering the extreme case where $E = \mathscr{E}$, it is clear that $\mathcal{M}/_{\mathscr{E}}$ is the finest partition the agent can have access to. Indeed, it is made of the finest equivalence classes that can ever be distinguished from a sensory input during the refinement process. From the agent's point of view, these sets of sensor poses are sensory equivalent but an outside viewer might distinguish them as different poses in space. These finest equivalence classes will be called *sensitive poses* in all the following. As shown later, they might be closely related to the notion of points in the physical space.

2) Illustrative example: All the aforementioned considerations were mainly theoretical. Let's now illustrate these points by using a very simple simulated robot agent made of one serial arm composed of two parts of identical length controlled by two revolute joints moving in a plane, see Figure 1. The end-effector of the system is endowed with a single-pixel camera which is only sensitive to illumination in such a manner that it can only send two values: $\boldsymbol{s} = 0$ if the illumination is zero and $\boldsymbol{s} = 1$ otherwise. The system is driven by two motor commands θ_1 and θ_2, which are supposed to represent directly the two joint angles, so that by convention $\theta_1, \theta_2 \in [-\pi, \pi[$ ($\theta_1 = \theta_2 = 0$ makes the arm horizontal). Suppose now that the environment is made of one black and one white areas separated by a straight line, as depicted on top of subfigures 1(a) and 1(b). Of course, the agent does not have access to this information and can only rely on its sensorimotor flow, i.e. variations of θ_1, θ_2 and their sensory consequences. At the very beginning, the set of all motor commands $\boldsymbol{m} = (\theta_1, \theta_2) \in \mathcal{M} = [-\pi, \pi[^2$ have not been distinguished from each other so that the current finest motor partition is $\{\mathcal{M}\}$. After having explored one black-and-white environment, the agent is able to obtain a finer motor partition. Indeed, two equivalent classes can easily be formed by regrouping all the motor commands $\boldsymbol{m}$ giving the same sensation, namely, for an environmental state ϵ, $[\boldsymbol{m}_0]_\epsilon$ for $s = 0$ and $[\boldsymbol{m}_1]_\epsilon$ for $s = 1$. Then, the set $\{[\boldsymbol{m}_0]_\epsilon, [\boldsymbol{m}_1]_\epsilon\}$ forms a partition of the set $\{\mathcal{M}\}$, which can be represented as two separated points, see Figure 1(d). This partition is also colored directly in the motor set in Figure 1(a). Of course this partition is environment dependent, which is captured in the formalization with a dependency to the environmental state ϵ in the equivalence relation $=_{\Psi_\epsilon}$. This dependency has been discussed in many publications [16], [17], [21], [26], and no clear solutions have been proposed so far to deal with this environment variability. In these works, the environment is systematically considered static, and they often restrict their study to cases where the environment changes do not influence the sensorimotor flow (by working on the agent body, for instance, like in [19] and [27]).

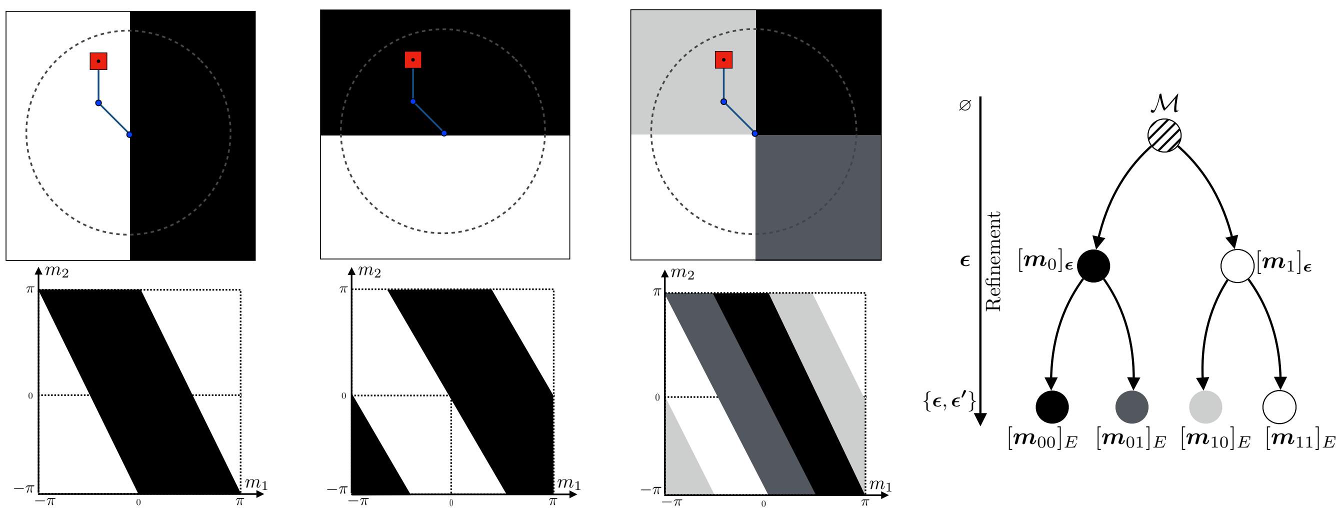
However if the environmental state changes to a new state ϵ' (corresponding to a new black-and-white separation of the robot's pose space, as shown in Figure 1(b)), then it is possible that previously inseparable motor configurations (regrouped in one equivalence class) are now generating different sensations. Considering this new environmental state ϵ' alone, it is clear that the agent can partition its motor set into two equivalence classes $[\boldsymbol{m}_0]_{\epsilon'}$ and $[\boldsymbol{m}_1]_{\epsilon'}$, thus leading to a new motor partition shown in Figure 1(b). Remembering the previous partition, the agent can now build a finer partition for having sequentially experimented the environmental states ϵ *and* ϵ'. The resulting *multi-environment* partitioning can be easily deduced in this case, and is shown in Figure 1(c). In this intuitive example, the agent is now able to separate the equivalence class $[\boldsymbol{m}_0]_\epsilon$, which relates to all the motor configurations giving the same 0 sensation value *for the environmental state* ϵ, into two new subsets that are denoted $[\boldsymbol{m}_{00}]_E = [\boldsymbol{m}_0]_\epsilon \cap [\boldsymbol{m}_0]_{\epsilon'}$ and $[\boldsymbol{m}_{01}]_E = [\boldsymbol{m}_0]_\epsilon \cap [\boldsymbol{m}_1]_{\epsilon'}$, with the set $E = \{\epsilon, \epsilon'\}$. $[\boldsymbol{m}_1]_\epsilon$ is also partitioned in two subsets $[\boldsymbol{m}_{10}]_E$ and $[\boldsymbol{m}_{11}]_E$. Following the colors used in Figure 1(c), one can then illustrate this refinement with the down arrows in Figure 1(d).

If the experiment is reproduced, then the multi-environment partition will again be refined, with all the equivalence classes being more and more partitioned into smaller subsets. The refined sets would then show a monotonically growing number of points along with the number of environmental states observed. In this example, there is an infinite number of environment states that the agent can interact with so that the number of points shall tend to infinity: all the equivalence classes can always be further partitioned with a new specific environment. However, by considering the case where the agent has interacted with all possible environmental states (in fact, it is not a reasonable consideration as the empirical version of the process may only give at most a countably infinite number of experiences) one obtains a case where the equivalence classes are not refinable and can be considered as points, called the *sensitive poses*. But what are these sensitive poses in the actual physical space?

B. *From the motor quotient set to the sensor pose*

So far, the previous section has highlighted the only two sets the agent can be aware of: the motor configuration set $\mathcal{M}$ and the sensory set $\mathcal{S}$, where its motor states and sensory inputs respectively lie. Both sets are linked together through the sensorimotor law Ψ unknown to the agent. From an external point of view, the sensory input $\boldsymbol{s} \in \mathcal{S}$ is generated by rigid sensors whose spatial state in the world can be entirely described by their *pose* in the world: $\boldsymbol{x} \in \mathcal{X}$, with $\mathcal{X}$ the *sensors pose set*. Let's focus on this new set and highlight the links between $\mathcal{X}$, $\mathcal{M}$ and $\mathcal{S}$.

1) Definition of the sensor pose set: It is well known in robotics that the *forward kinematics function* f, which accounts for the relative movements allowed at each joint, is dependent on the geometry of the robot, is a function linking the motor state $\boldsymbol{m}$ to the corresponding sensors pose $\boldsymbol{x}$, usually in Euclidean space, so that $\boldsymbol{x} = f(\boldsymbol{m})$. The pose set $\mathcal{X}$ is built such that f is surjective, i.e. all sensor poses can be obtained from a motor configuration in $\mathcal{M}$). In general, the pose $\boldsymbol{x}$—which is a parameterization of the sensors spatial state in the physical world—refers to the sensors positions



(a) Environment $\boldsymbol{\epsilon}$. (b) Environment $\boldsymbol{\epsilon'}$. (c) Sequence $(\boldsymbol{\epsilon},\boldsymbol{\epsilon'})$. (d) Motor set refinement.

Fig. 1. Illustration of the refinement process. (Top) The 2D serial agent with 2 degrees of freedom and the single-pixel camera (red) and its pose space (circle). (Bottom) Motor equivalence classes represented in the motor configuration set $\mathcal{M}$. (a) The agent is in an environment made of black and white areas parameterized by its state variable $\boldsymbol{\epsilon}$. The agent is only able to split its motor set in two parts, i.e. two equivalence classes $[\boldsymbol{m}_0]_{\boldsymbol{\epsilon}}$ and $[\boldsymbol{m}_1]_{\boldsymbol{\epsilon}}$ respectively related to the sensory inputs: $\boldsymbol{s} = 0$ (black) and $\boldsymbol{s} = 1$ (white). (b) The same applies for the new environment configuration $\boldsymbol{\epsilon'}$. (c) When remembering the environmental states $\{\boldsymbol{\epsilon}, \boldsymbol{\epsilon'}\}$, the agent obtains 4 equivalence classes represented with 4 shades of gray: black is for the indexed sensory set $\{(\epsilon, 0), (\epsilon', 0))\}$, dark gray is for $\{(\epsilon, 0), (\epsilon', 1))\}$, light gray is for $\{(\epsilon, 1), (\epsilon', 0))\}$ and white for $\{(\epsilon, 1), (\epsilon', 1))\}$. (d) Then, the initial two equivalence classes ($[\boldsymbol{m}_0]_{\boldsymbol{\epsilon}}$ and $[\boldsymbol{m}_1]_{\boldsymbol{\epsilon}}$) built when experiencing the environmental state $\boldsymbol{\epsilon}$ can be partitioned again in 4 subsets after experiencing a new state $\boldsymbol{\epsilon'}$.

and their orientations relatively to the frame of the agent's body. The sensors poses are externally defined thus not directly accessible by the agent. The sensory input is thus linked to the spatial state of the sensors through the *forward sensory function* $\phi_{\boldsymbol{\epsilon}}$, so that $\boldsymbol{s} = \phi_{\boldsymbol{\epsilon}}(\boldsymbol{x})$. In the end, the sensorimotor law $\Psi_{\boldsymbol{\epsilon}}$ can be written as the composition $\Psi_{\boldsymbol{\epsilon}} = \phi_{\boldsymbol{\epsilon}} \circ f$, which is summarized by the diagram

$$\begin{tikzcd} \mathcal{M} \arrow[r, "f"] \arrow[rr, bend left, "\Psi_{\boldsymbol{\epsilon}}"] & \mathcal{X} \arrow[r, "\phi_{\boldsymbol{\epsilon}}"] & \mathcal{S} \end{tikzcd} . \tag{7}$$

Introducing $\mathcal{X}$ is a convenient way to understand how the motor refinement, outlined in § II-A1, is related to space. Following the same ideas, one can define an equivalence relation for two poses. It is noteworthy that the two functions f and $\phi_{\boldsymbol{\epsilon}}$ are possibly non-injective. This means that two different motor configurations can lead to the same sensors pose (i.e. the non-injectivity of f captures the agent kinematics redundancy) and in a specific environment state $\boldsymbol{\epsilon}$ two different sensors poses can lead to the same sensory state (i.e. the non-injectivity of $\phi_{\boldsymbol{\epsilon}}$ captures the environmental redundancies but also the sensors possible symmetries). In the vein of Equation (2), for any $\boldsymbol{\epsilon} \in \mathscr{E}$, one can again define an equivalence relation $=_{\phi_{\boldsymbol{\epsilon}}}$ for two poses with

$$\boldsymbol{x}_1 =_{\phi_{\boldsymbol{\epsilon}}} \boldsymbol{x}_2 \Leftrightarrow \phi_{\boldsymbol{\epsilon}}(\boldsymbol{x}_1) = \phi_{\boldsymbol{\epsilon}}(\boldsymbol{x}_2). \tag{8}$$

Thus, one can regroup all the sensor poses leading to the same sensory state in their equivalence class

$$[\boldsymbol{x}]_{\boldsymbol{\epsilon}} = \{\boldsymbol{r} \in \mathcal{X}; \boldsymbol{r} =_{\phi_{\boldsymbol{\epsilon}}} \boldsymbol{x}\}. \tag{9}$$

Then, the quotient set $\mathcal{X}/_{\boldsymbol{\epsilon}} = \{[\boldsymbol{x}]_{\boldsymbol{\epsilon}}; \boldsymbol{x} \in \mathcal{X}\}$ forms a refinement of the trivial partition $\{\mathcal{X}\}$. By generalization over multiple environmental states, one can then define the *multi-environment* equivalence relation $=_{\phi_E}$ for any subset $E \subseteq \mathscr{E}$ defined as

$$\begin{aligned} \boldsymbol{x}_1 =_{\phi_E} \boldsymbol{x}_2 &\Leftrightarrow \phi_{\boldsymbol{\epsilon}}(\boldsymbol{x}_1) = \phi_{\boldsymbol{\epsilon}}(\boldsymbol{x}_2), \forall \boldsymbol{\epsilon} \in E \\ &\Leftrightarrow \phi_E(\boldsymbol{x}_1) = \phi_E(\boldsymbol{x}_2), \end{aligned} \tag{10}$$

where the function

$$\phi_E(\boldsymbol{x}) = \{(\epsilon, \phi_{\boldsymbol{\epsilon}}(\boldsymbol{x})); \boldsymbol{\epsilon} \in E\} \in \mathcal{S}_E \tag{11}$$

maps each sensor pose to its respective *indexed sensory set* acquired along the environmental states in E. The equivalence relation $=_{\phi_E}$ can be understood as: two poses are said equivalent after having seen all environments in $E \subseteq \mathscr{E}$ if the sensory inputs they have generated are equal for all environmental states in E. These poses can then be regrouped in an equivalence class

$$[\boldsymbol{x}]_E = \{\boldsymbol{r} \in \mathcal{X}; \boldsymbol{r} =_{\phi_E} \boldsymbol{x}\}. \tag{12}$$

The set of all equivalence classes is the quotient set $\mathcal{X}/_E$. Like before, the extreme case where $E = \mathscr{E}$ is of particular interest. Indeed, $\mathcal{X}/_{\mathscr{E}}$ is made of equivalence classes which can not be further fragmented into subsets, thus defining the *sensitive pose set*. It is interesting to see that the set of elements of $\mathcal{X}$ is in fact a refinement of $\mathcal{X}/_{\mathscr{E}}$. This highlights the fact that there might exists some subsets of positions and orientations in the sensor's pose set that can never be distinguished from sensory inputs. Then, from an internal point of view using the sensorimotor flow, the agent will never be capable to separate those ambiguous subsets and is unable to represent the whole pose set $\mathcal{X}$.

2) Interlink between $\mathcal{M}/_E$ and $\mathcal{X}/_E$: So far, two quotient sets have been introduced: (i) the motor quotient set $\mathcal{M}/_E$ which can be built directly from the sensorimotor flow and by interaction between the agent and its environment, and (ii) the quotient pose set $\mathcal{X}/_E$ which captures sensor poses that have not yet been distinguished from sensory inputs. For each point in $\mathcal{X}/_E$ corresponds a unique set of indexed sensations. The map f being surjective, all poses have been obtained from a motor configuration in $\mathcal{M}$. Therefore, the same set of indexed sensations has also been generated from a unique equivalence class in $\mathcal{M}/_E$. It then follows that there is a bijection between the sets $\mathcal{X}/_E$ and $\mathcal{M}/_E$ and the set of generated indexed sensations. Then, from an external point of view, the link between all the sets defined so far can be subsumed by the following diagram

$$\begin{tikzcd} \mathcal{M} \arrow[r, "f"] \arrow[rr, bend left, "\Psi_E"] \arrow[d, "\pi_E^{\mathcal{M}}"] & \mathcal{X} \arrow[r, "\phi_E"] \arrow[d, "\pi_E^{\mathcal{X}}"] & \mathcal{S}_E \\ \mathcal{M}/_E \arrow[r, "f/_E"] & \mathcal{X}/_E \arrow[ur, "\phi_E/_E"'] & \end{tikzcd}, \qquad (13)$$

where $f/_E$ represents the unique bijective map mapping together equivalence classes from $\mathcal{M}/_E$ to $\mathcal{X}/_E$. $\pi_E^{\mathcal{M}}$ and $\pi_E^{\mathcal{X}}$ both represent the canonical projections from points to equivalence classes. Consequently, the agent can exploit $\mathcal{M}/_E$ as an internal representation of $\mathcal{X}/_E$. Furthermore, letting $E = \mathscr{E}$ means then that the agent has experienced all the possible environmental states, i.e. each finest equivalence class in $\mathcal{X}/_{\mathscr{E}}$ is equivalently represented by a finest equivalence class in the internal representation $\mathcal{M}/_{\mathscr{E}}$. These considerations are illustrated in the following subsection.

3) Illustrative example (cont'd): Let's come back to the previous illustrative example, where a 2-DOF robot arm explores a black-and-white environment. In this simple case:

- the environmental state $\boldsymbol{\epsilon}$ can be described by a straight line delimiting the pose space in two areas together with a binary value indicating which one is black;
- the agent's motor configuration set $\mathcal{M}$ is made of the set of the two joint angles θ_1, θ_2 so that $\mathcal{M} = \{(\theta_1, \theta_2); \theta_1, \theta_2 \in [-\pi, \pi[\}$;
- the forward kinematics function f gives the end-effector position $(x, y) \in \mathcal{X} \subset \mathbb{R}^2$ as a function of θ_1, θ_2 with $x = L(\cos\theta_1 + \cos(\theta_1+\theta_2))$ and $y = L(\sin\theta_1 + \sin(\theta_1+\theta_2))$, where L is the length of both arm parts;
- the sensor, placed at (x, y), delivers a sensory input[2] $\boldsymbol{s} = \phi_{\boldsymbol{\epsilon}}(x, y) \in \mathcal{S} = \{0, 1\}$.

The agent is there endowed with a point sensor, so a pose $\boldsymbol{x}$ in $\mathcal{X}$ is nothing else but a point in a 2D Euclidean space. Since two distinct points in the 2D Euclidean plan can always be separated by a straight line, equivalently for two distinct poses in $\mathcal{X}$, there always exists an environmental state $\boldsymbol{\epsilon} \in \mathscr{E}$ for which the corresponding sensations are distinct. Then, for a given pose $\boldsymbol{x}$, the equivalent class $[\boldsymbol{x}]_{\mathscr{E}}$ regrouping all the poses that always give the same sensations for all $\boldsymbol{\epsilon} \in \mathscr{E}$ is just the singleton $[\boldsymbol{x}]_{\mathscr{E}} = \{\boldsymbol{x}\}$. This means that the finest partition $\mathcal{X}/_{\mathscr{E}}$ of the pose set is the set of points in $\mathcal{X}$, i.e. $\mathcal{X}/_{\mathscr{E}} = \{\{(x, y)\}, (x, y) \in \mathcal{X}\}$.

[2]In the particular case where the sensor is placed exactly on the straight line splitting the working space in two areas, it is arbitrary chosen that $\boldsymbol{s} = 0$.

Following the same ideas, it is clear that the equivalence classes in the motor configurations set $\mathcal{M}$ are the set of motor configurations leading to a same and unique pose through the forward kinematics function f. Consequently, the finest partition $\mathcal{M}/_{\mathscr{E}}$ is made of equivalence classes $[\boldsymbol{m}]_{\mathscr{E}}$ individually corresponding to one equivalence class $[\boldsymbol{x}]_{\mathscr{E}} = [f(\boldsymbol{m})]_{\mathscr{E}}$. However, the finest equivalence classes in $\mathcal{M}/_{\mathscr{E}}$ have a unique corresponding point in $\mathcal{X}/_{\mathscr{E}}$, which have been shown to represent *points* in the 2D Euclidean pose space. Then, without knowledge on the forward kinematics function f—and through a refinement strategy—the agent can build the set $\mathcal{M}/_{\mathscr{E}}$ which captures kinematics redundancies and constitutes a very good candidate for representing the actual space of Euclidean poses.

All these considerations are represented in Figure 2. The pose space is represented in the left, where each pose can be reached by the agent from one or multiple motor configurations due to the kinematics redundancy. For instance, the pose $\boldsymbol{x}_1$ (resp. $\boldsymbol{x}_3$) can be reached by the 2 different motor configurations $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ (resp. $\boldsymbol{m}_3$ and $\boldsymbol{m}_4$) in $\mathcal{M}$. The same applies for the pose $\boldsymbol{x}_5$ located at the limit of the pose, which can be obtained with a unique motor configuration $\boldsymbol{m}_5$. Another particular case is the pose $\boldsymbol{x}_6$ obtained when the sensor is exactly in the center of the pose space, which can be reached with all motor configurations $\boldsymbol{m} = (\theta_1, \theta_2)$ such that $\theta_2 = -\pi$, thus building the set $\mathscr{M}_6 \in \mathcal{M}$. As explained above, each of these poses is linked to an equivalent class in $\mathcal{M}/_{\mathscr{E}}$ once the agent has experienced all the possible environmental states $\boldsymbol{\epsilon}$ in $\mathscr{E}$. For instance, the two motor configurations $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ (resp. $\boldsymbol{m}_3$ and $\boldsymbol{m}_4$) can be regrouped in the equivalent class $[\boldsymbol{m}_1]_{\mathscr{E}} = \{\boldsymbol{m}_1, \boldsymbol{m}_2\}$ (resp. $[\boldsymbol{m}_3]_{\mathscr{E}} = \{\boldsymbol{m}_3, \boldsymbol{m}_4\}$) in $\mathcal{M}/_{\mathscr{E}}$. Then, one can see on this illustration that each indivisible equivalent class obtained on the finest partition $\mathcal{M}/_{\mathscr{E}}$ forms a sensitive pose, each of them being associated to a unique pose in the pose space, i.e. a point in the 2D Euclidean space. Thus, the agent knows for instance that any motor configuration selected in $\mathscr{M}_6$ will correspond to a sensitive pose point $[\boldsymbol{m}_6]_{\mathscr{E}}$ (with $\boldsymbol{m}_6 \in \mathscr{M}_6$) and so to a unique pose in the sensitive pose set. In that sense, one can qualitatively understand that the finest refinement represents the kernels of the forward kinematics function.

From the mathematical formalism previously proposed, one seems to have definitely concluded on the way an agent can refine its motor configurations from the set of sensory inputs generated during the environment exploration. However, the equivalence classes have been defined very intuitively by introducing a somewhat sequential exploration of the environmental states. Moreover the environment is possibly continuous with an uncountable number of environmental states and the agent can not experience all of them. From the agent's point of view, the environmental states it can experience can been considered as a statistical sample of the set of all environmental states. In order to generalize the agent experience one needs to change the current deterministic framework and introduce statistical properties on the set of environmental states. Indeed, if some sensory distinction between two motor configurations have a zero probability to occur, then these configurations must be considered equivalent from the agent point of view. This allows us to extend the formalism to *observable/unobservable*

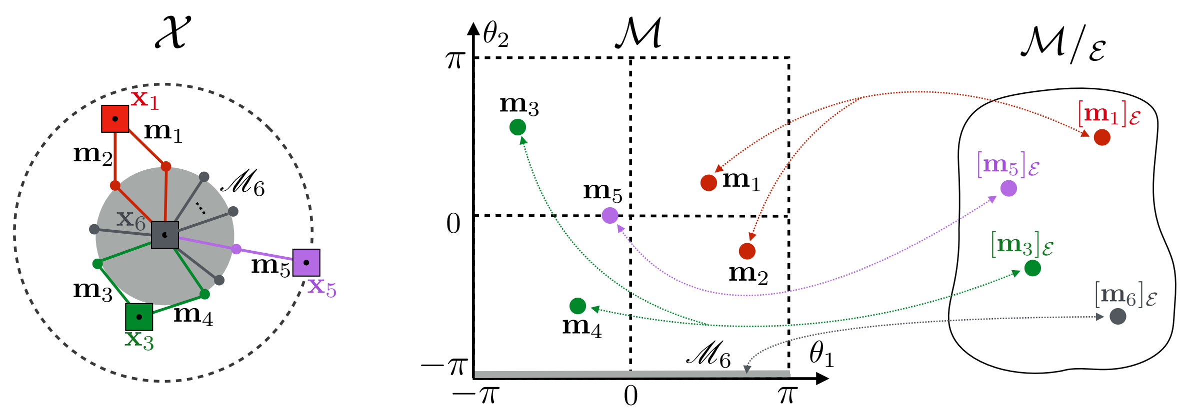


Fig. 2. Illustration of the link between $\mathcal{X}$, $\mathcal{M}$ and $\mathcal{M}/_{\mathscr{E}}$ for the simple example used in the paper. (Left) Pose set, with the single-pixel camera represented as a square. (Middle) Motor configuration set. (Right) Sensorimotor representative set, i.e. the finest partition of the motor set. In the end, and for the finest motor representation, each finest equivalence class $[\boldsymbol{m}_i]_{\mathscr{E}}$ in $\mathcal{M}/_{\mathscr{E}}$ represents only one *point* in the agent pose space $\mathcal{X}$.

events using the probability and measure theory.

C. On the observability of the finest refinement

1) Formalism: Because the number of possible environmental states is possibly uncountable, one needs to properly define the probability space of interest (see [28] as a reference on probability and measure). Let's take $\mathscr{E}$ as the sample space and its power set $\mathcal{P}(\mathscr{E})$ as the σ-field or set of events. $\mathcal{P}(\mathscr{E})$ is composed of all possible subsets of environmental states the agent can interact with. Let's assume that there exists a probability measure P from a subset of environmental states $E \in \mathcal{P}(\mathscr{E})$ to $[0,1]$ that represents the probability for the agent to interact with an environmental state inside E. If a subset E has a probability measure $P(E) > 0$ then it is called *observable*, meaning that there is nonzero probability to observe at least one element of it. Therefore, one could derive a new definition of the finest sensory equivalence classes on the basis on this probability measure. Indeed, let's consider two poses and E the set of environmental states that make them generate distinct sensations. E might be empty, finite or even infinite, but if the measure $P(E) = 0$, then the probability to observe an environmental state that distinguish both poses is 0. In other terms, the sensory distinguishability of such poses is unobservable by the agent. Then, one can extend the concept of sensitive poses to *observable* sensitive poses. From the probability measure P on the set of environmental states $\mathscr{E}$, lets $=_P$ be the new equivalence relation such that, for any pair $\boldsymbol{x}_i$, $\boldsymbol{x}_j \in \mathcal{X}$ of poses,

$$\boldsymbol{x}_i =_P \boldsymbol{x}_j \Leftrightarrow P(\{\epsilon \in \mathscr{E}; \boldsymbol{x}_i \neq_{\phi_\epsilon} \boldsymbol{x}_j\}) = 0. \tag{14}$$

This equivalence relation can be interpreted as: in order for two poses to be in the same equivalence class, it suffices that the probability to observe an environmental state that separates them is zero. Let's now denote by $\mathcal{X}/_P = \{[\boldsymbol{x}]_P; x \in \mathcal{X}\}$ the *observable sensitive pose set* made of the equivalence classes given by the equivalence relation $=_P$. From an external point of view, $\mathcal{X}/_P$ is the finest set of points that can be observed from the sensorimotor flow; it is thus all that can be represented by the agent. Obviously, one can also define the equivalence relation on the set of motor configurations to obtain the agent internal representation

$$\boldsymbol{m}_i =_P \boldsymbol{m}_j \Leftrightarrow P(\{\epsilon \in \mathscr{E}; \boldsymbol{m}_i \neq_{\Psi_\epsilon} \boldsymbol{m}_j\}) = 0. \tag{15}$$

And define $\mathcal{M}/_P = \{[\boldsymbol{m}]_P; \boldsymbol{m} \in \mathcal{M}\}$ as the internal representation. With the same considerations than in §II-B2, $\mathcal{M}/_P$ has the same number of points than in $\mathcal{X}/_P$. All these properties can be subsumed in the following diagram

$$\begin{array}{ccc} \mathcal{M} & \xrightarrow{f} & \mathcal{X} \\ {\scriptstyle \pi_P^{\mathcal{M}}}\big\downarrow & & \big\downarrow{\scriptstyle \pi_P^{\mathcal{X}}} \\ \mathcal{M}/_P & \xrightarrow{f/_P} & \mathcal{X}/_P \end{array} \tag{16}$$

where $f/_P$ denotes the bijection that maps the equivalence classes together, i.e. $f/_P : [\boldsymbol{m}]_P \rightarrow [f(\boldsymbol{m})]_P$ for any $\boldsymbol{m} \in \mathcal{M}$. $\pi_P^{\mathcal{M}}$ and $\pi_P^{\mathcal{X}}$ both represents the two canonical projections from points to equivalence classes.

2) Illustrative example (cont'd): Let's apply these new considerations to the previous example. Considering Figure 2, if one takes two separated poses in the pose space and chooses at random a straight line intersecting with the pose space, the probability to have a sensory difference between them is different from 0. Indeed, if the distribution of straight lines intersecting with the pose space is uniform, there is non-null probability to separate two distinct points with this straight line. Therefore, the sensitive pose set $\mathcal{X}/_{\mathscr{E}}$ and the observable sensitive pose set $\mathcal{X}/_P$ are identical. Moreover, the probability measure P qualitatively gives a notion of distance between poses, the farther they are, the higher is the probability to distinguish them because the subset of environmental states that separates them is bigger.

The next section investigates how the probability measure P gives a structural information on the internal representation and how it is related to the structure of the physical space.

III. Introducing Structural Considerations into the Mathematical Formalism

So far, it has only been shown that a refinement process allows the agent to obtain a set of points $\mathcal{M}/_P$ in bijection with the set of observable sensitive poses in the set $\mathcal{X}/_P$. Until now, the representation is correct if the internal representation captures the observable sensitive points. However, in order to be exploitable, the agent should also be able to represent the continuity of movements in the physical space. Yet, there is no guarantee that close points in $\mathcal{X}/_P$ are represented by close points in $\mathcal{M}/_P$. These considerations are carefully addressed in the following subsections.

A. Sensorimotor structures on the quotient spaces

From a mathematical point of view, $\mathcal{X}/_P$ and $\mathcal{M}/_P$ have been described as *sets*, and not *spaces*. Transforming these two sets into spaces requires the introduction of some additional *structure*. In fact, it is possible to derive an intrinsic dissimilarity measure between the points of these sets as they can be linked together by the probability measure P on the set $\mathscr{E}$. Indeed, let's consider any pair of poses $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{X}$. Then from the probability measure P, one can derive the probability $p(\boldsymbol{x}_i, \boldsymbol{x}_j)$ for the event *"experience an environment generating two different sensations at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$"*. This can be formalized by splitting up the set of environmental states $\mathscr{E}$ in two complementary subsets E_{ij} and E_{ij}^c, with E_{ij} the set of environmental states such that $\forall \boldsymbol{\epsilon} \in E_{ij}$, $\phi_{\boldsymbol{\epsilon}}(\boldsymbol{x}_i) \neq \phi_{\boldsymbol{\epsilon}}(\boldsymbol{x}_j)$. Then, using P, the aforementioned probability is given by $p(\boldsymbol{x}_i, \boldsymbol{x}_j) = P(E_{ij})$. Consequently, if $p(\boldsymbol{x}_i, \boldsymbol{x}_j) = 0$ then the poses are equivalent and $\boldsymbol{x}_i =_P \boldsymbol{x}_j$. Importantly, it is proven in Appendix A that p is a pseudometric on $\mathcal{X}$. Because p is a pseudometric, distinct points in $\mathcal{X}$ may have a distance p equal to 0. However, in the quotient set $\mathcal{X}/_P$ these points are reduced in the same equivalence classes. Therefore, the pseudometric p induces a metric p^* on the quotient set $\mathcal{X}/_P$ defined as $p^*([\boldsymbol{x}_i]_P, [\boldsymbol{x}_j]_P) = p(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Equivalently, one can define the probability σ for pairs of motor configurations $\boldsymbol{m}_i, \boldsymbol{m}_j \in \mathcal{M}$ to observe an environmental state that generates different sensations with $\sigma(\boldsymbol{m}_i, \boldsymbol{m}_j) = p(f(\boldsymbol{m}_i), f(\boldsymbol{m}_j)) = P(E_{ij})$. σ is called the *sensory dissimilarity*, it is also a pseudometric on $\mathcal{M}$ and it induces a metric σ^* on the quotient set $\mathcal{M}/_P$ defined as $\sigma^*([\boldsymbol{m}_i]_P, [\boldsymbol{m}_j]_P) = \sigma(\boldsymbol{m}_i, \boldsymbol{m}_j)$. Therefore, it appears the agent can actually build a metric *space* $(\mathcal{M}/_P, \sigma^*)$ by exploiting the refinement process highlighted in the previous section. Moreover, the bijective map

$$(\mathcal{M}/_P, \sigma^*) \xrightarrow{f/_P} (\mathcal{X}/_P, p^*) \tag{17}$$

has the following property: $\sigma^*([\boldsymbol{m}_i]_P, [\boldsymbol{m}_j]_P) = p^*(f/_P([\boldsymbol{m}_i]_P), f/_P([\boldsymbol{m}_j]_P))$. Therefore it is an isometry and both spaces are linked together by a distance-preserving transformation.

One has to keep in mind that the metric structures induced by the probability measure P are inherited from the statistics of sensorimotor invariants. Consequently, they are *empirical* because built directly from the observable sensorimotor *experience*, i.e. from the comparison of sensations along the agent life. These empirical structures, as defined previously, are not arbitrary but are an intrinsic property of the sensorimotor experience of the agent. Based on previous considerations, the space $(\mathcal{M}/_P, \sigma^*)$ is metrically equivalent to the space $(\mathcal{X}/_P, p^*)$, however, does this empirical structure respect the continuity of the physical world? If not, the agent should not have any interest in building $(\mathcal{M}/_P, \sigma^*)$ as it would not constitute a good representation of the world. This specific point is addressed in the next subsection.

B. Natural topology: towards a good representation respecting continuity of the physical world

Before dealing with the notion of a "good representation of the world", one needs to introduce *natural* structures in the considered spaces. Here, "*natural*" refers to the structures which are directly induced by intrinsic properties of the physical world in which the agent is embedded such as continuity. Indeed, the displacements of the agent sensors are hypothesized as being continuous in time and space. Therefore one can endow respectively the motor set $\mathcal{M}$ and pose set $\mathcal{X}$ with topological structures $\tau_{\mathcal{M}}$ and $\tau_{\mathcal{X}}$ such that $\tau_{\mathcal{M}}$ guarantees the continuity of action and $\tau_{\mathcal{X}}$ the continuity of the forward kinematics function $f(.)$, with $f : (\mathcal{M}, \tau_{\mathcal{M}}) \to (\mathcal{X}, \tau_{\mathcal{X}})$.

So, it appears that the quotient pose set $\mathcal{X}/_P$ can be endowed with two different topological structures:

- the one induced by the quotient of $(\mathcal{X}, \tau_{\mathcal{X}})$ (i.e. the quotient topology), which captures its *natural* topology;
- the one induced by the metric p^*, which is *empirically* obtained by the agent through sensorimotor experience.

Since one wishes the empirical topology to represent the continuity of the physical world, both topological structures must be equivalent. Under the following two hypotheses, it is proven in Appendix B that these two structures on the quotient set $\mathcal{X}/_P$ are indeed topologically equivalent.

(H1) The probability p is a continuous property of the physical space, or equivalently p is a continuous map from $\mathcal{X} \times \mathcal{X}$ with the product topology to $\mathbb{R}_{\geq 0}$.
(H2) The agent's motor configuration space $(\mathcal{M}, \tau_{\mathcal{M}})$ is compact.

Additionally, since relation (17) states that $\mathcal{M}/_P$ and $\mathcal{X}/_P$ with their empirical structures are homeomorphic, then the quotient motor space $\mathcal{M}/_P$ with the empirical structure is topologically equivalent with the space $\mathcal{X}/_P$ endowed with the natural structure. Thus, $(\mathcal{M}/_P, \sigma^*)$ can be considered to be a good topological representation of $\mathcal{X}/_P$ respecting the physical continuity.

Now that we have stated the hypotheses under which the agent might be interested in building $\mathcal{M}/_P$, let's focus in the next section on how to perform the refinement process from an experimental point of view.

IV. An experimental framework for the refinement process

The notion of refinement, together with considerations on the possible structures inherited from the agent's actions and statistics of its sensorimotor invariants, have been introduced in the two previous section. Yet, all these points were mainly theoretical: no considerations on the computational process the agent should undergo in order to obtain a correct representation have been underlined. However, the refinement is a process that will be conducted during the agent life, i.e. along time. This section is devoted to the introduction of such experimental considerations that should allow to formalize (i) what structures can actually be represented and how, and (ii) how an outside viewer can experimentally assess if this internal representation is correct. These two points are addressed in the next two subsections.

A. *Introducing an experimental point of view on the refinement process*

1) Experimental setup: To begin with, let's consider that the agent is totally naive and only has access to its uninterpreted sensorimotor flow. From a known home state $\boldsymbol{m}_0$, the agent performs a naive babbling through a set of N randomly generated—but repeatable—actions, and then goes back to its home state $\boldsymbol{m}_0$. Through this exploration, the agent obtains a number of N reads of its sensorimotor flow $(\boldsymbol{s}_i, \boldsymbol{m}_i)$, with $i = 1, \ldots, N$. Let's call $M = \{\boldsymbol{m}_i\}_i$ the *motor exploration set* corresponding to the reached motor configurations. Under the hypothesis of repeatability of actions, the agent can repeat the exploration of the set M, each repetition being parameterized by an integer k. Thus, at repetition k and motor configuration $\boldsymbol{m}_i$, the agent sensory input is noted $\boldsymbol{s}_i[k]$.

At the end of each repetition of M, the agent can compare the sensations between all pairs $(\boldsymbol{m}_i, \boldsymbol{m}_j)$ of motor configurations in the exploration set M. For repetition k, one can define the $N \times N$ *dissimilarity matrix* $\mathbf{D}[k]$ of all comparisons whose elements $D_{ij}[k]$ are computed as

$$D_{ij}[k] = \frac{1}{k}\sum_{l=1}^{k} \delta_{ij}[l], \text{ with } \delta_{ij}[l] = \begin{cases} 0 \text{ if } \boldsymbol{s}_i[l] = \boldsymbol{s}_j[l], \\ 1 \text{ otherwise.} \end{cases} \tag{18}$$

The elements $D_{ij}[k]$ of the dissimilarity matrix $\mathbf{D}[k]$ represent the frequency for two motor configurations $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$ of being separated by a sensory input, during the sample of k repetitions. Thus, it can be envisaged as an estimator of the sensory dissimilarity $\sigma(\boldsymbol{m}_i, \boldsymbol{m}_j)$ defined in §III-A.

2) Experimental internal representation: From Equation (18) one has that, if during each repetition from 1 to k the environmental states can be considered static, then the zeros in $\mathbf{D}[k]$ actually represent the equivalence classes mentioned in the formalization in §II for an explored set M and a sequence of k environmental states. Thus, at the beginning of the agent life, i.e. when $k = 0$, $\mathbf{D}[0]$ can be initialized as a null matrix: all motor configurations have not been distinguished from each other (see the initial partition in the top of Figure 1(d)). Then, as the number of repetitions increases, the agent can notice that during some repetitions two sensations $\boldsymbol{s}_i[k]$ and $\boldsymbol{s}_j[k]$ might differ. Then, for such pairs one has $D_{ij}[k] > 0$ and the agent motor configurations set M can be partitioned as illustrated in the graph of Figure 1(d).

One would like to state that when k tends to infinity the dissimilarity D_{ij} converges in probability towards $\sigma(\boldsymbol{m}_i, \boldsymbol{m}_j)$, but it needs some additional theoretical requirements on both the probability measure P and the stochastic process describing the succession of environmental states. Moreover the case where the environment is fixed during a repetition is purely theoretical. Dealing with a real-life scenario requires to take into account a continuously changing environment. These changes might cause distortions between the dissimilarity D_{ij} and the theoretical dissimilarity $\sigma(\boldsymbol{m}_i, \boldsymbol{m}_j)$. Furthermore, as the agent does not explore all its motor configurations space, it cannot represent its entire pose quotient space with the natural structure. The dissimilarity can rather be interpreted as an estimation of the metric structure of the space where the agent sensors have moved. Considerations about the link between the structures arising from the refinement process in a realistic environment are developed in the following subsection.

B. *Evaluation of the representation*

So far, it has been proven that, under some hypotheses, the agent is *theoretically* able to capture topological properties of the quotient pose space. In the experimental case, it is not possible to assess the convergence of the refinement process at a topological level because the explored points are discrete, and the discrete topology is trivial. The evaluation based on topological continuity cannot be performed and should be replaced by the evaluation of preservation of local structures in the form of "small" neighborhoods. Points that are close in one space should correspond to close points in the other space. Indeed it has been shown that the internal representation is a finite metric space which is equivalent to a fully connected, undirected, weighted graph where points are nodes, and edges are weighted by the distance between the two linked points. Until now, the quotient pose space has only been given a topological structure. In order to empirically evaluate the internal representation metric, it is mandatory to define a metric on the quotient pose set. This will allow a proper evaluation of the metric distortion of the representation.

1) Evaluation metric and local structure in the represented space: Let's assume that the represented space, e.g. the quotient pose space $\mathcal{X}/_P$, is endowed with a metric ρ, called the *evaluation metric*, known to an external viewer and compatible with the natural topological structure in the quotient pose space $\mathcal{X}/_P$. This metric could be used in traditional applications to derive cost functions for tasks such as path planning. In the experimental case, when the agent explores the set of motor configurations M, it runs through the discrete set $X = f(M)$ of poses. After taking the quotient by regrouping points that are theoretically not distinguishable from a sensory point of view, we obtain the discrete subset $X/_P$ of quotient pose set $\mathcal{X}/_P$ for which ρ is also a metric. Therefore, the space to be represented by the agent is the discrete set $X/_P$ with the distance matrix $\mathbf{R}$ whose elements R_{ij} corresponds to the distances between elements in $X/_P$: $\rho([\mathbf{x}_i]_P, [\mathbf{x}_j]_P)$. The discrete metric space $(X/_P, \mathbf{R})$ can also be represented as a weighted graph.

2) Evaluation criteria: From one side we have the internal representation $(M, \mathbf{D})$ which evolves with the time of exploration, and on the other side the represented space $(X/_P, \mathbf{R})$. We propose two useful evaluation criterion to evaluate the structural similarity between these two spaces. The first criterion will guaranty that the agent has distinguished all points that can theoretically be distinguished. The second criterion, gives an evaluation on the conservation of the local structure between the represented and the representative graphs.

a) The refinement criterion C_1: this criterion is defined as the ratio of the pairs of configurations $(\boldsymbol{m}_i, \boldsymbol{m}_j)$ that have not been distinguished by a sensory difference yet but are distinct in the quotient pose space, over all the N^2 possible pairs. C_1 can then be computed as

$$C_1 = \frac{|\{(\boldsymbol{m}_i, \boldsymbol{m}_j); D_{ij} = 0 \text{ and } R_{ij} \neq 0\}|}{N^2}. \tag{19}$$

The finest refinement is then obtained when $C_1 = 0$, meaning that all distinct points in the quotient pose space are distinct in the internal representation.

b) Local structure similarity criterion C_2: topological preservation between discrete models such as graphs or SOMs is usually dealt with by evaluating the preservation of neighborhoods [29]. Here it is evaluated as a measure of how well small neighborhoods are preserved between the dissimilarity $\mathbf{D}$ and the evaluation metric $\mathbf{R}$ and is based on the *adjusted Locally Continuous Meta-Criteria (LCMC)* used for Local Multi-Dimensional Scaling (LMDS) [30]. The adjusted LCMC is mainly used in the context of nonlinear multidimensional reduction. It evaluates both the preservation of continuity and the trustfulness of an embedding from measures of dissimilarity on a dataset to a low dimensional Euclidean space. The choice of adjusted LC meta-criteria is also justified by the fact that it is a non-metric criterion, as it uses ranks in the dissimilarities and not metric information, and so is invariant to monotonous scaling of the dissimilarities and the evaluation metric.

The adjusted LCMC is computed as follows. Let's $\mathcal{N}_K^{\mathbf{D}}(i) = \{j_1, \cdots, j_K\}$ be the K-Nearest Neighbors (K-NNs) of configuration i with regard to dissimilarity matrix $\mathbf{D}$, and $\mathcal{N}_K^{\mathbf{R}}(i) = \{k_1, \cdots, k_K\}$ the K-NNs with regard to the evaluation metric $\mathbf{R}$. Then the neighborhood similarity for point i is simply the cardinality of their common K-NNs:

$$N_K(i) = |(\mathcal{N}_K^{\mathbf{D}}(i) \cap \mathcal{N}_K^{\mathbf{R}}(i))|. \tag{20}$$

The adjusted LCMC is given in its global form by a normalized and adjusted average over all points by

$$Q(K) = \frac{1}{KN} \sum_{i=1}^{N} N_K(i) - \frac{K}{N-1}. \tag{21}$$

A value of $Q(K)$ close to 1 indicates a high similarity between all the K-NNs in both spaces. However the adjusted LCMC is a function of the number of nearest neighbors: the higher K, the bigger the considered neighborhoods. But the only interest of the criterion C_2 is to evaluate similarity on "local structures" associated to a small value of K. This can be achieved thanks to the approach by Lee et al. [31], which consists in finding the value K_{max} of K which maximizes the adjusted LCMC along

$$K_{\text{max}} = \text{argmax } Q(K). \tag{22}$$

Then, criterion C_2 is computed as the average below K_{max} of Q as

$$C_2 = \frac{1}{K_{\text{max}}} \sum_{K=1}^{K_{\text{max}}} Q(K). \tag{23}$$

The quantity C_2 assesses the similarity of local structures, e.g. the preservation of neighbors insides the K_{max}-NNs. K_{max} represents the scale that corresponds to "local" considerations. The value $C_2 = 1$ indicates that all the neighborhoods of all points of size inferior and equal to K_{max} are perfectly preserved. The value $C_2 = 0$ can be interpreted as a random permutation of the points.

3) 2D visualization: Additionally to the two criterion introduced previously, it is also possible to project the dissimilarity matrix $\mathbf{D}$ to a low-dimensional Euclidean space by using multidimensional scaling (MDS). This projection can then be visualized to assess qualitatively the resemblance between the projected internal representation and the actual quotient pose space. However, this visualization can not replace the two quantitative criterion C_1 and C_2 in the general case, since the quotient pose space cannot always be embedded into a 2D or 3D Euclidean space without big distortions. The algorithm used for the projection is Isomap [32]. Given a value K, Isomap performs Classical MDS using geodesic distances on the K-NN graph of the dissimilarity matrix $\mathbf{D}$. The visualization being not a criterion *per se*, any non-metric multidimensional scaling algorithm that preserves local structures such as Local-MDS, SOM, LLE, tSNE or Curvilinear Component Analysis could have been selected. The choice of Isomap is based on its simplicity and the fact that the neighborhood scale K is already available from the computation of the local similarity criterion K_{max} in Equation (22).

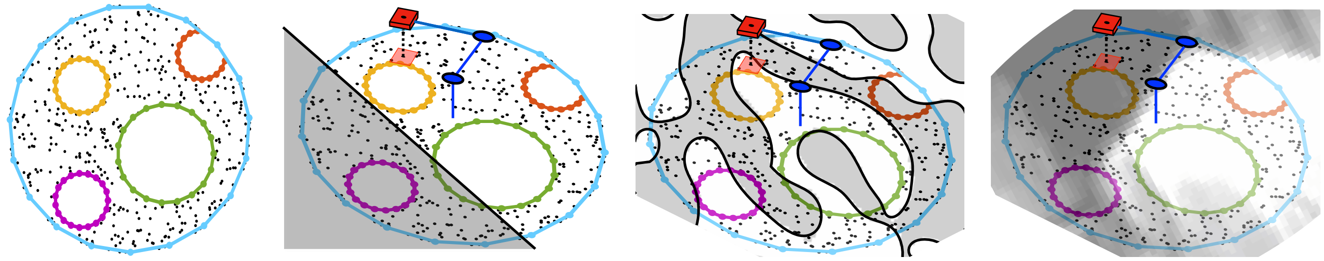
The process for building the internal representation has now been formalized. Two criteria have been introduced to evaluate if the space $(M, \mathbf{D})$ is a good representative of local structures in $(X/_P, \mathbf{R})$. The next section shows the results of the implementation of the refinement process for a simulated agent in different environments.

V. Simulated Results and Discussion

This section aims at providing a proof of concept on how a naive agent can build an internal representation of its working space in an unknown environment with an uninterpreted sensorimotor flow. Therefore, the simulated agent will be tested on different environments. The detailed simulation setup is described in a first subsection. Then, quantitative and qualitative evaluations of the obtained representations for two different scenarios of increasing complexity are proposed.

A. Simulation setup

The agent used in the simulations is the simple 2 degrees of freedom agent introduced in §II-A2. For all scenarios, the agent's motor exploration set M is chosen in the following way. First, as written in the illustrative example in §II-B3, recall that the motor configurations of the agent is represented with the tuple $\boldsymbol{m} = (\theta_1, \theta_2)$ where θ_1 and θ_2 are the two joint angles of the serial arm. The agent starts from a home position $\boldsymbol{m}_0 = (0, 0)$, the sensor being at the far right of the working space. It then generates random actions simulated by an addition of two random angles sampled from a uniform probability distribution in $[0, 2\pi]$ after which it goes back to its home position $\boldsymbol{m}_0$. After the i-th action, the agent receives the motor



(a) Original working space with structural clues (b) (Straight lines) environment (c) (Blobs) environment (d) (Movie) environment

Fig. 3. Far left: the 2D agent's working space after exploring a motor configuration set M with $N = 500$ configurations. The explored sensor poses are black dots and the structural clues are linked colored dots. Right: the three different background environments.

configuration $\boldsymbol{m}_i$. Without further exploration, the explored pose space would be a set of random points inside a 2D disk. These explored poses are the black points in figure 3(a). For a better visual interpretation, some structural clues have been added in the form of forced exploration points, shown as linked colored points in figure 3(a). Then the explored points inside these rings have been removed from the exploration set, thus allowing a better visualization of the internal representation distortion. At the end, the explored motor configuration space M is composed of $N = 500$ motor configurations and sensor poses for every scenario. At the end of repetition k, the agent computes the dissimilarity distance $\mathbf{D}[k]$ with equation (18) between all pairs of explored configurations. For the simulated environments, the space to be represented is the pose space $X/_P = X = f(M)$, where f is the forward kinematics of the agent, and the evaluation metric is the euclidean distance between the poses $R_{ij} = \rho(\boldsymbol{x_i}, \boldsymbol{x_j}) = ||\boldsymbol{x}_i - \boldsymbol{x}_j||_2$.

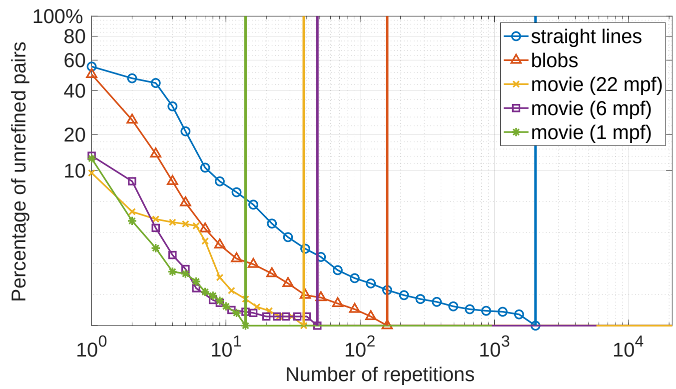


Fig. 4. Evolution of the refinement criterion C_1 for the 5 different environments. The vertical lines show at which repetition the finest refinement ($C_1 = 0$) is obtained.

B. Environments description and results

The simulated environments are separated in 2 scenarios of increasing complexity. All environments are black-and-white or gray-scale backgrounds on the pose space of the 2D agent as shown in the three subfigures 3(b), 3(c) and 3(d). The first scenario is composed of 2 different environments. In this first scenario the refinement process matches with the theory: the environmental states are kept fixed during the exploration of M, the sensory inputs are either 0 or 1. This corresponds to the theoretical set-up formalized in §II. The second scenario is more realistic and is composed of 3 different environments. In this second scenario the environmental states are allowed to change during the exploration as it would happen in a realistic environment, and the agent's sensory inputs can take more values than 0 or 1.

1) Scenario 1, static environment during exploration:

a) Straight lines environment: In this first environment, the environmental states are identical to those in the illustrative example and are depicted in subfigure 3(b). Each environmental state is randomly chosen as a straight line crossing the pose space separating the background in one black and one white areas. The distribution of these straight lines is taken so that they uniformly cover the working space, see method 2 of Bertand's Paradox [33]. The agent's sensory input is either 1 or 0 according to which side of the straight line the agent's sensor is in. At the end of each repetition, a new straight line is randomly chosen and the refinement process continues. This process is repeated until $k = 10^6$ explorations of the motor exploration set.

Figure 4 plots the evolution of the refinement criterion C_1. The finest refinement is obtained after almost 1400 repetitions, meaning that the environment is very slow at separating points in the representation. It is indeed very structured by opposition to random sensory values over the working space. In Figure 6 is shown the evolution of the local similarity criterion C_2 between the measured dissimilarity $\mathbf{D}[k]$ at repetition k and the Euclidean distance $\mathbf{R}$ between the poses of the target pose space shown in subfigure 3(a). The criterion C_2 starts very low for the first exploration and then converges towards a value of $C_2 = 0.98$ which is almost a perfect match of local neighborhoods. Indeed, for this environment the statistics of sensory invariants are invariant to translations or rotations in the 2D Euclidean working space [34] and in fact it can be proved that the measured dissimilarity converges in probability to the Euclidean distance between the poses, up to a constant factor. In order to interpret the values of C_2 and C_1, let's visualize the 2D projection of the obtained dissimilarity matrix along the agent life using Isomap. The results of the visualization for the different environments are

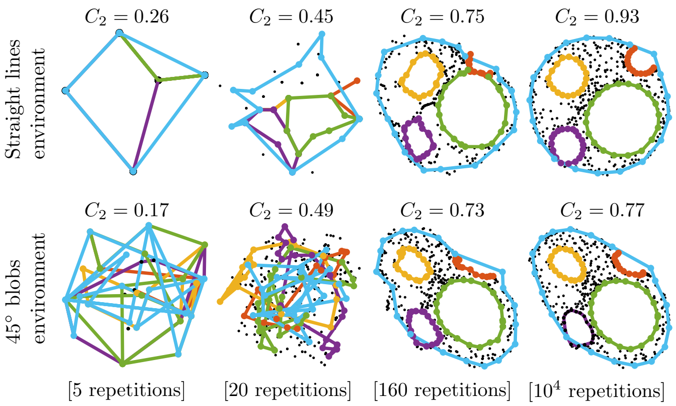


Fig. 5. Visualization of criterion C_2 with the corresponding internal representations projected in 2D using Isomap with $K_{\max}$-NNs during the agent's life in the static environment scenario.

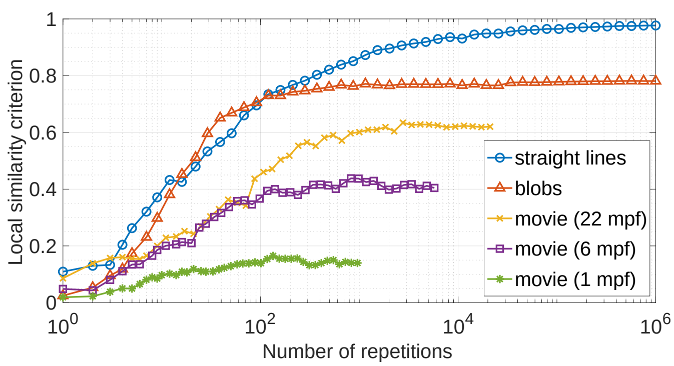


Fig. 6. Evolution of the local similarity criterion C_2. (straight lines) converges to $C_2 = 0.98$, (blobs) converges to $C_2 = 0.78$ and (movies 22 mpf) approximately converges to $C_2 = 0.62$, (movies 6 mpf) approximately converges to $C_2 = 0.40$ and (movies 1 mpf) converges towards $C_2 = 0.14$. The closer C_2 is to 1 the more accurate is the local structure of the internal representation.

shown in Figure 5 and 7[3]. The internal representation starts with a few number of distinct points. After 20 repetitions, the number of points in the representation has increased, and the structural cues starts shaping with $C_2 = 0.4$. They are visually well preserved starting from $C_2 = 0.6$ and 50 repetitions. After the almost convergence of C_2 around 10^4 repetitions, the internal representation is visually a quasi-perfect reproduction of the working space which is confirmed by a C_2 close to 1.

b) Blobs environment: In this second environment, the environmental states are composed of randomly generated background images with black and white blobs stretched in the bottom left/top right direction of the working space, as plot in figure 3(c). The blobs comes from a random noise generated using a procedural Perlin noise with anisotropic filtering (steerable Gaussian filter) oriented at 45 degrees. The resultant noisy image is then thresholded to give black and white blobs. The agent sensations are either 1 or 0 depending on the color of the blob the sensor is looking at. After each repetition, a new black-and-white image is randomly sampled.

[3]Five video attachments have been uploaded together with this submission. They show the evolution of the representation for the interaction of the agent with the 5 environments.

This process is repeated until $k = 10^6$ explorations of the motor exploration set.

Looking at the refinement criterion C_1 represented in Figure 4, one can see that the finest refinement is obtained about ten times quicker than for the straight lines environment. Indeed, far poses have a clear tendency to be separated quicker than in the previous environment since blobs are more localized in space. For a given agent, one could then compare environments with respect to the number of repetitions required to reach the finest refinement: the smaller this number, the richer the environment. The richness of an environment is thus defined as its ability to quickly separate pairs of poses of an agent's explored pose set. In Figure 6, the local similarity criterion C_2 converges to a value of $C_2 = 0.77$ indicating a distortion of small dissimilarities in $\mathbf{D}$ with respect to the evaluation metric $\mathbf{R}$. Indeed, the sensory invariants have a higher probability to occur for a pair of points aligned along the top left/bottom right direction because the blobs are stretched in this direction. Thus, these pairs of points have a lower dissimilarity and are considered closer than in the orthogonal direction. The projected internal representation in Figure 5 shows as expected a stretch in the direction of high sensory invariants variance. However, after convergence, the structural cues are visually well preserved indicated by a high $C_2 = 0.77$, but not as well as in the previous environment.

Until now, the environmental states were kept fixed during the exploration. This gives us an insight into the representation distortion as well as into the notion of richness of the environment. Let's now consider more realistic dynamical environments.

2) Scenario 2, dynamical environment (Movie): In this last scenario, the environmental states are composed of cropped images of a black-and-white movie frames. One cropped image is shown in Figure 3(d). The agent hovers its sensor across the image; for that purpose, the agent has been centered inside the movie frame and scaled so that the diameter of its working space corresponds to 40 pixels in the image. In order to simulate a spatially continuous environment, the value given by its sensor is the linearly interpolated gray value at the $2D$ sensor pose from adjacent pixels which is quantified on a 0 to 15 gray scale. Of course the agent's sensory inputs are uninterpreted: it cannot know that a value of 3 is closer to 4 than to 15. Each sensory value is actually seen in this framework as a symbol. To simulate the dynamics of the environment, the image is refreshed with the next frame of the movie after a given number of movements. In the environments 3, 4 and 5 the agent respectively moves at a speed of 22 motor configurations per frame (mpf), 6 mpf and 1 mpf: thus, a high mpf indicates low dynamics of the environment. The movie file[4] has been played 3 times, resulting in 479166 frames, which in turns corresponds to 21082 repetitions for environment 3, 5749 repetitions for environment 4 and 960 repetitions for environment 5. Repeating the same movie allows for more samples with the same sensory statistics so that convergence is obtained slightly quicker than when playing differents movies.

[4]The selected movie is Phantom of the Opera (1943).

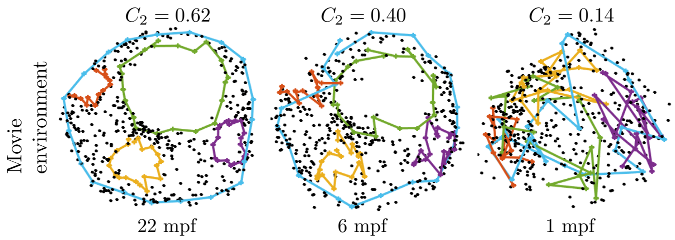


Fig. 7. Final values of C_2 with the projected internal representations for 3 different dynamics (mpf = movements per frame) of the movie environments. The 3 final representations are obtained after 10^4, 5700 and 960 repetitions respectively.

Criterion C_1 plot in Figure 4 shows that the refinement is quicker in scenario 2 which indicates richer environments. The first reason comes from a higher number of sensory values: 16 gray scale values instead of 0 and 1, making it less probable to have sensory invariants. The 1 mpf environment reaches the finest refinement the quickest, indicating that quick refinement is also caused by a higher relative dynamics of sensory changes in the environment with respect to agent's movements. Moreover, Figure 6 exhibits that the 22 mpf, 6 mpf and 1 mpf environments respectively converges to $C_2 = 0.62$, $C_2 = 0.4$ and $C_2 = 0.14$. In these dynamical environments, a pair of close points in space but explored after a long time might have a distortion caused by the possible environment changes during the reaching time, thus the higher the dynamics the higher the distortion after convergence. At the bottom of Figure 7 are shown the final internal representations obtained after convergence of C_2. Visually the structural clues are well represented when dealing with slow relative dynamics (22 mpf and 6 mpf movie environments). However, considering a high relative dynamics of 1 mpf, the agent's movements are too slow with respect to changes in the environment. As a consequence, two close poses in space that have a high probability to generate the same sensory input if reached quickly, have now a smaller probability to produce the same sensory input as the environment is likely to change during the agent movement. The agent cannot capture the spatial order of the poses anymore and the representation is almost a random permutation of points, which is obtained when $C_2 = 0$ by definition. In this last case, the internal representation is certainly not exploitable for any task defined in the working space because the agent cannot plan continuous trajectories. Note that the proposed approach could also be exploited in real world environments. The main difficulties would be to deal with repeatability of the agent movements, i.e. control problems such as possible collisions or variability of the actions. One can also mention some possible issues regarding the existence of noise in the sensory inputs. In the end, all this should result in a more distorted representation; hopefully, there is room for increasing the robustness of the approach, for instance by using reinforcement learning [35].

VI. Conclusion

This paper dealt with the question of how and why a totally naive (interpretation free) agent with access to its actuator states and sensors inputs can, through active exploration, build an internal representation of its sensors states in the physical world. It has been shown that, under the assumption of continuity on the statistics of sensory invariants in the physical space, this internal representation is topologically equivalent to a space called the quotient pose space. The quotient pose space represents the states of the agents that can't be distinguished by a sensory input (with probability 1). This space comes with a "natural topology" defined as the finest one that makes the movement of the sensors continuous. Then a formalization has been proposed to adapt the process of refinement to an experimental context with realistic environments. Proof of concept examples are shown in adequate environments: low dynamics relatively to the agent's movements, high probability of sensory invariants, but some limitations are presented when dealing with more challenging environments. A step forward would be to show a direct exploitation of the topological internal representation obtained after the refinement process in tasks such as path planning or obstacle avoidance possibly in real world applications. Moreover, the introduced concept of sensorimotor structure learning could be supplemented with the introduction of hierarchical structures, and the proposed formalism could be used as a grounding for higher level considerations in interactive perception. It may also be used as a tool for learning the structure of space based on the compensatory actions in Poincaré's intuition.

Appendix A
Proof that p is a pseudometric in the pose space

In order to show that p is a pseudometric in the pose space $\mathcal{X}$, let us prove that it satisfies the pseudometric conditions: for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{X}$:

(a) $p(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ (non-negativity).
(b) $p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y}, \boldsymbol{x})$ (symmetry).
(c) $p(\boldsymbol{x}, \boldsymbol{z}) \leq p(\boldsymbol{x}, \boldsymbol{y}) + p(\boldsymbol{y}, \boldsymbol{z})$ (triangle inequality).

Proof:

a) Non-negativity: P being a probability measure, it is non-negative, e.g. for any $E \subseteq \mathscr{E}, P(E) \geq 0$. Hence $p(\boldsymbol{x}, \boldsymbol{y}) = P(\{\epsilon \in \mathscr{E}; \phi_\epsilon(\boldsymbol{x}) \neq \phi_\epsilon(\boldsymbol{y})\}) \geq 0$ so p is non-negative.

b) Symmetry: $p(\boldsymbol{x}, \boldsymbol{y}) = P(\{\epsilon \in \mathscr{E}; \phi_\epsilon(\boldsymbol{x}) \neq \phi_\epsilon(\boldsymbol{y})\}) = P(\{\epsilon \in \mathscr{E}; \phi_\epsilon(\boldsymbol{y}) \neq \phi_\epsilon(\boldsymbol{x})\}) = p(\boldsymbol{y}, \boldsymbol{x})$.

c) Triangle inequality: Let's define $E_{\boldsymbol{xy}}$ as the set $\{\epsilon \in \mathscr{E}; \phi_\epsilon(\boldsymbol{x}) \neq \phi_\epsilon(\boldsymbol{y})\}$, and equivalently for $E_{\boldsymbol{yz}}$ and $E_{\boldsymbol{xz}}$. Then $p(\boldsymbol{x}, \boldsymbol{y}) + p(\boldsymbol{y}, \boldsymbol{z}) = P(E_{\boldsymbol{xy}}) + P(E_{\boldsymbol{yz}})$. P being a probability measure it is subadditive so that $P(E_{\boldsymbol{xy}} \cup E_{\boldsymbol{yz}}) \leq P(E_{\boldsymbol{xy}}) + P(E_{\boldsymbol{yz}})$. Let us show that $E_{\boldsymbol{xz}} \subseteq E_{\boldsymbol{xy}} \cup E_{\boldsymbol{yz}}$ by contradiction. Assume that $E_{\boldsymbol{xz}} \not\subseteq E_{\boldsymbol{xy}} \cup E_{\boldsymbol{yz}}$, then there exists at least one $\epsilon \in E_{\boldsymbol{xz}}$ which is also outside of $E_{\boldsymbol{xy}} \cup E_{\boldsymbol{yz}}$. Then $\phi_\epsilon(\boldsymbol{x}) \neq \phi_\epsilon(\boldsymbol{z})$ by definition of $E_{\boldsymbol{xz}}$ but because ϵ is neither in $E_{\boldsymbol{xy}}$ nor in $E_{\boldsymbol{yz}}$, necessarily $\phi_\epsilon(\boldsymbol{x}) = \phi_\epsilon(\boldsymbol{y})$ and $\phi_\epsilon(\boldsymbol{y}) = \phi_\epsilon(\boldsymbol{z})$, so $\phi_\epsilon(\boldsymbol{x}) = \phi_\epsilon(\boldsymbol{z})$ which gives a contradiction. Therefore by monotonicity of the probability measure, $P(E_{\boldsymbol{xz}}) \leq P(E_{\boldsymbol{xy}} \cup E_{\boldsymbol{yz}}) \leq P(E_{\boldsymbol{xy}}) + P(E_{\boldsymbol{yz}})$, hence $p(\boldsymbol{x}, \boldsymbol{z}) \leq p(\boldsymbol{x}, \boldsymbol{y}) + p(\boldsymbol{y}, \boldsymbol{z})$.

Appendix B
Proof that empirical and natural topologies on the quotient set are topologically equivalent.

Let us denote as $\tau_{\mathcal{M}}$, $\tau_{\mathcal{X}}$ and $\tau_{\mathcal{X}/P}$ the natural topologies of the respective spaces $\mathcal{M}$, $\mathcal{X}$ and $\mathcal{X}/P$. Let's call τ_p the topology induced by the pseudometric p in space $\mathcal{X}$ and τ_{p^*} for the topology induced by the metric p^* in space $\mathcal{X}/P$. Let's consider the following commutative diagram:

$$\begin{tikzcd} (\mathcal{M}, \tau_{\mathcal{M}}) \arrow[r, "f"] & (\mathcal{X}, \tau_{\mathcal{X}}) \arrow[r, "\mathrm{id}_{\mathcal{X}}"] \arrow[d, "\pi_P^{\mathcal{X}}"'] & (\mathcal{X}, \tau_p) \arrow[d, "\pi_P^{\mathcal{X}}"] \\ & (\mathcal{X}/P, \tau_{\mathcal{X}/P}) \arrow[r, "\mathrm{id}_{\mathcal{X}/P}"] & (\mathcal{X}/P, \tau_{p^*}) \end{tikzcd} \tag{24}$$

From hypothesis **(H1)** the pseudometric p is continuous on the pose space $\mathcal{X}$ with the natural topology, hence the identity map $\mathrm{id}_{\mathcal{X}}$ from $(\mathcal{X}, \tau_{\mathcal{X}})$ to $(\mathcal{X}, \tau_p)$ is continuous. Moreover, $\pi_P^{\mathcal{X}}$ which maps poses to equivalence classes is a continuous map from $(\mathcal{X}, \tau_{\mathcal{X}})$ to $(\mathcal{X}/P, \tau_{\mathcal{X}/P})$ because, from the definition of natural topology, $\tau_{\mathcal{X}/P}$ is the topology induced by this mapping. The map $\pi_P^{\mathcal{X}}$ from the pseudometric space $(\mathcal{X}, \tau_p)$ to the metric space $(\mathcal{X}/P, \tau_{p^*})$ is called a *metric identification* and it preserves the induced topologies, thus it is continuous. Hence, by commutation of the diagram, the map $\mathrm{id}_{\mathcal{X}/P}$ from $(\mathcal{X}/P, \tau_{\mathcal{X}/P})$ to $(\mathcal{X}/P, \tau_{p^*})$ is also a continuous map. From **(H2)**, we have that $(\mathcal{M}, \tau_{\mathcal{M}})$ is a compact space. Compactness property is preserved through continuous maps so that all considered spaces are compact. It has been shown that the space $(\mathcal{X}/P, \tau_{p^*})$ is a metric space, in particular it is Hausdorff (all pairs of distinct points have disjoint neighborhoods). Moreover, it is known that a continuous bijection from a compact space to a Hausdorff space is a homeomorphism so that $\mathrm{id}_{\mathcal{X}/P}$ is a homeomorphism. Therefore, natural and empirical topologies coincide on $\mathcal{X}/P$, $\tau_{\mathcal{X}/P} = \tau_{p^*}$.

Acknowledgment

The authors would like to thank Jean-Merwan Godon for his advice on the theoretical developments written in this paper.

References

[1] M. Charlet, E. Marcellini, and C. Gosselin, "Trajectory planning of projectile catching maneuvers for robotic manipulators," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 08 2018, p. V05BT07A014.

[2] Y. Zhao, X. Chai, F. Gao, and C. Qi, "Obstacle avoidance and motion planning scheme for a hexapod robot octopus-iii," *Robotics and Autonomous Systems*, vol. 103, pp. 199 – 212, 2018.

[3] D. Shaikh, L. Bodenhagen, and P. Manoonpong, "Concurrent intramodal learning enhances multisensory responses of symmetric crossmodal learning in robotic audio-visual tracking," *Cognitive Systems Research*, vol. 54, pp. 138 – 153, 2019.

[4] J. K. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness." *The Behavioral and brain sciences*, vol. 24, no. 5, pp. 939–973; discussion 973–1031, Oct. 2001.

[5] J. K. O'Regan, "Solving the "real" mysteries of visual perception: The world as an outside memory," *Canadian Journal of Psychology*, vol. 46, no. 3, pp. 461–488, 1992.

[6] 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.

[7] G. Le Clec'H, J. K. O'Regan, and B. Gas, "Acquisition of a space representation by a naive agent from sensorimotor invariance and proprioceptive compensation," *International Journal of Advanced Robotic Systems*, vol. 13, no. 6, 2016.

[8] H. Poincaré, "L'espace Et la Géométrie," *Revue de Métaphysique Et de Morale*, vol. 3, no. 6, pp. 631–646, 1895.

[9] ——, "Science and hypothesis," *London: Walter Scott Publishing*, 1905.

[10] J. J. Gibson, *The senses considered as perceptual systems*. Boston: Houghton, 1966.

[11] P. Bach-y Rita, C. Collins, F. Saunders, B. White, and L. Scasdden, "Vision substitution by tactile image projection," *Nature*, vol. 221, no. 5184, pp. 963–964, mar 1969.

[12] J. J. Gibson, *The ecological approach to visual perception*. Boston: Houghton, 1979.

[13] A. Berthoz, *Le sens du mouvement*. Odile Jacob, 1997.

[14] A. Noë, *Action in perception*. The MIT Press, 2004.

[15] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme, "Interactive perception: Leveraging action in perception and perception in action," *IEEE Transactions on Robotics*, vol. PP, no. 99, pp. 1–19, 2017.

[16] 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.

[17] 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.

[18] P. Demartines and J. Herault, "Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets." *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 148–54, jan 1997.

[19] 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.

[20] A. Laflaquiere, "Grounding the experience of a visual field through sensorimotor contingencies," *Neurocomputing*, 2017.

[21] V. Y. Roschin, A. Frolov, Y. Burnod, and M. Maier, "A neural network model for the acquisition of a spatial body scheme through sensorimotor interaction," *Neural Computation*, vol. 23, pp. 1821–1834, 2011.

[22] C. R. Shalizi and J. P. Crutchfield, "Computational mechanics: Pattern and prediction, structure and simplicity," *Journal of Statistical Physics*, vol. 104, pp. 816–879, 2001.

[23] P. Capdepuy, D. Polani, and C. L. Nehaniv, "Constructing the basic umwelt of artificial agents: An information-theoretic approach," in *Advances in Artificial Life*, F. Almeida e Costa, L. M. Rocha, E. Costa, I. Harvey, and A. Coutinho, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 375–383.

[24] D. Pierce and B. J. Kuipers, "Map learning with uninterpreted sensors and effectors," *Artificial Intelligence*, vol. 92, no. 1, pp. 169 – 227, 1997.

[25] L. Ollsson, C. Nehaniv, and D. Polani, "From unknown sensors and actuators to actions grounded in sensorimotor perceptions," *Connection Science*, vol. 28, no. 2, pp. 121–144, 2006.

[26] D. Philipona, J. K. O'Regan, J.-P. Nadal, and O. J.-M. Coenen, "Perception of the structure of the physical world using unknown sensors and effectors," *Advances in Neural Information Processing Systems*, vol. 15, 2004.

[27] A.A.Frolov, "Physiological basis of 3-D external space perception: approach of Henri Poincaré," in *History of the neurosciences in France and Russia*, hermann ed., 2011.

[28] P. Billingsley, *Probability and Measure*, 3rd ed. Wiley, 1995.

[29] T. Villmann, R. Der, M. Herrmann, and T. M. Martinetz, "Topology preservation in self-organizing feature maps: exact definition and measurement," *IEEE Transactions on Neural Networks*, vol. 8, no. 2, pp. 256–266, March 1997.

[30] L. Chen and A. Buja, "Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis," *Journal of the American Statistical Association*, vol. 104, no. 485, pp. 209–219, 2009.

[31] J. A. Lee and M. Verleysen, "Scale-independent quality criteria for dimensionality reduction," *Pattern Recognition Letters*, vol. 31, pp. 2248 – 2257, 2010.

[32] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[33] J. Bertrand, "Calcul des probabilités," p. 5, 1889.

[34] E. T. Jaynes, "The well-posed problem," *Foundation of Physics*, vol. 3, pp. 477–493, 1973.

[35] R. Jonschkowski and O. Brock, "Learning state representations with robotic priors," *Autonomous Robots*, vol. 39, no. 3, pp. 407–428, Oct 2015.