On the Sensory Commutativity of Action Sequences for Embodied Agents

Hugo Caselles-Dupré[1,2], Michael Garcia-Ortiz[2], David Filliat[1]
[1]Flowers Laboratory (ENSTA Paris & INRIA), [2]AI Lab (Softbank Robotics Europe)
caselles@ensta.fr, mgarciaortiz@softbankrobotics.com, david.filliat@ensta.fr

***Abstract*— We study perception in the scenario of an embodied agent equipped with first-person sensors and a continuous motor space with multiple degrees of freedom. We consider theoretically the commutation properties of action sequences with respect to sensory information perceived by such embodied agent. From the theoretical derivations, we introduce the Sensory Commutativity Probability criterion which measures how much an agent's degree of freedom affects the environment in embodied scenarios. We show how to compute this criterion in two environments, including a realistic robotic setup. We empirically illustrate how it can be used to improve sample-efficiency in Reinforcement Learning.**

I. Introduction

Perception is the medium by which agents organize and interpret sensory stimuli, in order to reason and act in an environment using their available actions [8]. We focus on scenarios where embodied agents are situated in *realistic* environments, i.e. the agents face partial observability, coherent physics, first-person view with high-dimensional state space and low-level continuous motor (i.e. action) space with multiple degrees of freedom. These embodied agents, when acting in such environment, produce a stream of sensorimotor data, composed of successions of motor states and sensory information. While most current approaches for building perception focus on studying the sensory information alone, several approaches [3, 9, 5, 22] that can be traced back to 1895 [18], advocate the necessity of studying the relation between sensors and motors for the emergence of perception.

Inspired by these work, we study the commutativity of action sequences with respect to sensors, which we term sensory commutativity, illustrated in Fig.1. We define the Sensory Commutativity Probability (SCP) as the probability that a sequence of movements using only one degree of freedom of the agent, an arm joint for instance, sensory commutes. We show that this value has meaning for the embodied agent: if the SCP is high then the degree-of-freedom has a low impact on the environment (e.g. moving a shoulder is more likely to lead to environment changes than moving a finger, so SCP for shoulder is lower than for finger). By computing the SCP for each degree of freedom of the agent, we are able to characterize its motor space and use this information for subsequent tasks. We illustrate this in our experiments as we show how SCP can be used to improve sample-efficiency in a Reinforcement Learning problem.

II. Related work

SensoriMotor theory (SMT) is a theory of perception that gives prominence to the role of motor information in the emergence of perceptive capabilities [15]. Inspired by philosophical ideas formulated more than a century ago by H.Poincare [18], it led to theoretical results regarding the extraction of the dimension of space [10], the characterization of displacements as compensable sensory variations [21], the grounding of the concept of point of view in the motor space [11, 12], as well as the characterization of the metric structure of space via sensorimotor invariants [13].

An important aspect of this literature is that action and sensor spaces have a shared underlying structure, since they are causally linked (sensory changes are caused by actions). It is suggested that the group structure would be well adapted [17, 18], yet it has never been formalized in these work. However recently, Symmetry-Based Disentangled Representation Learning (SBDRL) [6, 3] used group theory to formalize disentanglement in Representation Learning using symmetries, i.e. transformations of the environment that leave some aspects of it unchanged. Groups are composed of these transformations, and group actions are the effect of the transformations on the state of the world and representation. Inspired by this approach, we formalize the group structure suggested in the SMT theory and use it to define the SCP criterion.

In this paper we build on those previous works by choosing to study the set of action sequences, termed $Seq(\mathcal{M})$, and their commutative properties. We study the group and subgroup properties of $Seq(\mathcal{M})$, with the aim of organizing the motor space $\mathcal{M}$ hierarchically. This will be achieved with the definition of the Sensory Commutativity Probability criterion.

III. Formalism choice

We propose a mathematical framework for the embodied scenario which will allow to properly construct the Sensory Commutativity Probability. We start from the formalism used in SMT, which formalizes the perception of the agent as follows: $s_t = \phi(m_t, \epsilon_t)$.

At time t, the agent is in a particular motor state m_t, corresponding to the information about all the actionable parts of its body (joints, motors). We define everything external to the agent as the environment, itself in a state ϵ_t, e.g. a room with 6 walls plus light sources and objects placed in different locations. The agent can sense the world through its sensors s_t, and builds its perception by learning the sensorimotor

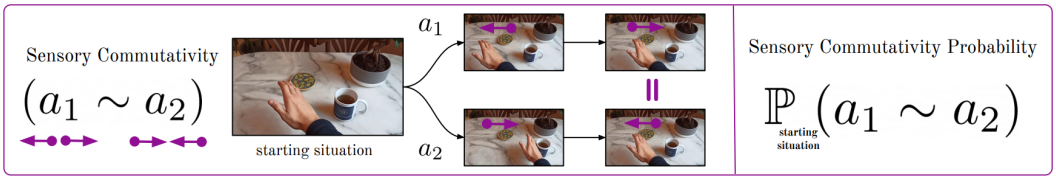

Figure 1: Two action sequences sensory commute if they produce the same sensory state when composed in different orders from the same starting position. In this example, the action sequences would not commute if an object would be in the way.

dependencies ϕ: a function that takes as input m_t and ϵ_t and produces sensory inputs from its sensors s_t. The agent can operate motor commands ($\Delta(m_t, m_{t+1})$), and the environment can change also through its own dynamics outside of the agent, represented by $\Delta(\epsilon_t, \epsilon_{t+1})$.

The dynamics of the world are generally not described in SMT, so we extend its formulation: $m'_{t+1}, \epsilon'_{t+1} = f(m_t, \epsilon_t, \Delta(m_t, m_{t+1}), \Delta(\epsilon_t, \epsilon_{t+1}))$.

IV. Properties of the set of action sequences $Seq(\mathcal{M})$

We will now attempt to formalize groups and sub-groups of symmetries. We propose G to be the set of motor command (or action) sequences of finite length, referred to as $Seq(\mathcal{M})$, and will attempt at extracting sub-groups based on subsets of these transformations.

A. Group structure of $Seq(\mathcal{M})$

[17] first defined a relation between action sequences: $h \sim g$ if and only if h and g affect the sensors in the same way. Using our formalism, we can translate this concept into an equality.

Definition 1. *Let $(h, g) \in Seq(\mathcal{M})$. h is equivalent to g under (m_t, ϵ_t), noted $h \sim_{m_t,\epsilon_t} g$ if and only if*

$$\phi(f(m_t, \epsilon_t, h, \Delta_{\epsilon_t}^{\epsilon_{t+1}})) = \phi(f(m_t, \epsilon_t, g, \Delta_{\epsilon_t}^{\epsilon_{t+1}}))$$

Intuitively, two actions sequences are equivalent for a particular motor state and environment state if applying them lead to the same sensory state. For instance for a multiple-joints arm moving freely in an empty space, there are multiple different ways of moving the arm from one motor state to another. This yields action sequences $(h_1, .., h_n)$ which are equivalent in this situation (m_t, ϵ_t), we thus have $h \sim_{m_t,\epsilon_t} g$. However in other situations these actions sequences can become not equivalent.

For convenience and clarity, we will drop the notation for dependence on (m_t, ϵ_t) and thus write $h \sim g$ whenever there are no ambiguities in the context. We now consider the structure of $Seq(\mathcal{M})$ under composition $\circ$ with respect to the equivalence $\sim$.

Proposition 1 (Structure of $(Seq(\mathcal{M}), \sim, \circ)$)**.**

1. $\sim$ is an equivalence, i.e. it is reflexive, transitive and symmetric.
2. $(Seq(\mathcal{M}), \circ)$ is a group w.r.t $\sim$.
3. $\circ$ is generally not commutative with respect to $\sim$.

Proof:

Proof is provided in Appendix A. ■

$(Seq(\mathcal{M}), \circ)$ is thus a group w.r.t $\sim$. This structure is consistent with the intuitions in SBRL and SMT theories. In the following, we build on the observation that composing action sequences is not generally commutative. We show how this property can lead the agent to organize and interpret its motor space.

B. Commutativity properties of $Seq(\mathcal{M})$

1) Philipona's conjecture: Philipona [17] already studied how action sequences commute with respect to the sensory information received by the agent. Action sequences do not necessarily commute as stated in Prop.1. For example if a movable object is placed to the right of your arm, moving your arm right then left will not have the same effect (in terms of sensor change) as moving it left then right. He conjectured that all action sequences that are not displacements commute with any action sequences. For instance moving you arms (displacement action) then opening the eyes (non-displacement action) will always commute whereas two displacement actions will not necessarily commute, depending on which starting situation (m_t, ϵ_t) is selected.

Conjecture 1 (Philipona's conjecture)**.** *The subset of $Seq(\mathcal{M})$ composed of non-displacements action sequences is the sub-group of $Seq(\mathcal{M})$ that commutes, i.e. the abelian sub-group of $Seq(\mathcal{M})$.*

We will illustrate this conjecture with experiments in Sec.V-B.

2) Sensory commutativity probability of an action: Based on Philipona's conjecture, we derive a criterion for characterizing how much each degree of freedom of the agent affects the world, computable using only sensorimotor data. We define "degree of freedom" (DOF) as a dimension of the multidimensional continuous action space of the agent.

Using the conjecture, we have that for an action sequence h, if the agent plays it in two different orders starting from the same situation, there is a chance that the agent will experience two different sensory outcomes only if the action sequence h is composed of at least one displacement action (an action that affect the environment such as moving limbs or going forward).

However not all displacement actions are equivalent. The agent is more likely to observe two different outcomes if the action sequence is composed of displacement actions that affect

the environment *a lot*. Consider moving your forearm (elbow joint) compared to moving your whole arm (shoulder joint): the latter is more likely to move things around in the environment and thus induce sensory non-commutativity when played in two different orders (i.e. having two different sensory outcomes). An elbow joint should therefore have a higher SCP than a shoulder joint.

We formalize this intuition by defining the Sensory Commutativity Probability (SCP) of a degree of freedom, averaged over all starting situations (m_t, ϵ_t):

Definition 2 (Sensory commutativity probability of a degree of freedom)**.** *Let $Seq(\mathcal{M}_k)$ be the set of motor commands (or action) sequences of finite length for the k^{th} degree of freedom of $\mathcal{M}$ (motor state space). Let $h \in Seq(\mathcal{M}_k)$ and let h_p be a random permutation of h (same sequence but different order).*

The Sensory Commutativity Probability of the k^{th} degree of freedom $SCP(\mathcal{M}_k)$ is defined as:

$$SCP(\mathcal{M}_k) = \mathbb{P}_{m_t,\epsilon_t,h}[h \sim_{m_t,\epsilon_t} h_p]$$

3) Sensory Commutativity Probability computation: We propose a simple procedure to estimate the SCP of each degree of freedom of the agent. We initialize the SCP value to 0 (`SCP`$\leftarrow$`0`). We then repeat the following process n times for each DOF:

- Sample an action sequence using the selected degree of freedom (a sequence of action where each action is a value between -1 and 1).

- Play it in 2 different orders starting from the same randomly chosen state and save the two final sensor images s_1 and s_2. Compute the distance between the two images $d(s_1, s_2)$.

- Count one (`SCP+=1`) if $d(s_1, s_2) \leq t$, zero otherwise.

Finally, the estimator of the SCP is the average over the number of trials (`SCP`$/n$). The parameters of the algorithm are the selected distance d that allows to compare the agent's observations, the threshold t and the number of iterations n. Note that using a simulation allows to play the two action sequences of different orders from the exact same starting position.

V. Sensory Commutativity Probability Experimental Analysis

In this first experimental section we compute and interpret the SCP for an embodied agent scenario. We then compare SCP to baseline alternatives. The simulation we use needs to satisfy the properties of an embodied agent scenario: navigable space with objects to interact with, first-person high dimensional observations, low-level high-dimensional action space and coherent physics. We start with a 2D environment and then move on to a more complex 3D realistic simulation.

A. Experimental setup

Simulation description.

Our first experiment uses Flatland [2], a platform for creating 2D RL environments. We construct an agent called Polyphemus (a Cyclop from the Greek mythology), that has a movable and rotatable base equipped with a rotatable head and two 2-DOF arms. The agent sees through its unique eye that has an activable eyelid, for a total of 8 DOF. The observation received by the agent is a 64 pixels RGB image. This agent is placed in a room with fixed, moving or movable entities, all of different colors. The agent can move around and interact with these entities. Its point of view can change through base movement, rotation, and head rotation. Our simulation is illustrated in Fig.2. For each degree of freedom, an action or motor command corresponds to a change in the longitudinal/angular velocity of the degree of freedom.

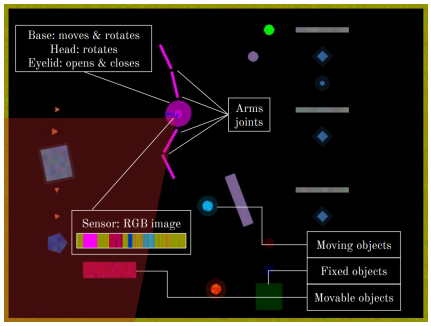

Figure 2: Simulation used for our experiments. The agent Polyphemus has a 8 DOF motor space, receives an image of it's only eye, and is placed in a room with fixed, movable and moving elements.

SCP computation. In order to compute the SCP of each of the 8 agent's degree of freedom, we have to select a distance and threshold as mentioned in Sec.IV-B3. The distance selected here is simply the mean squared error between s_1 and s_2, and the threshold is 0. This means that we consider that two action sequences sensory commutes if and only if applying the two action sequences from the same initial state lead to exactly the same sensors. We also experiment with two baselines described in App.B.

B. Results

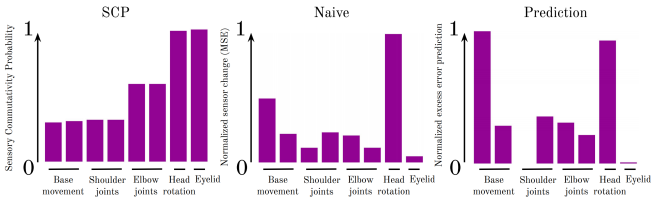

Figure 3: **Left:** Sensory Commutativity Probability for each degree of freedom. **Middle:** Naive alternative. **Right:** Prediction error alternative.

The results are consistent with Philipona's conjecture. Fig.3 (Left) shows that only two actions have an SCP of 1: *eyelid* and *head rotation*. All other actions have a SCP inferior to 1. This is consistent with Philipona's conjecture (Sec.IV-B1): *eyelid* and *head rotation* are the two degrees

of freedom that are **not** associated to displacements, thus action sequences composed of actions of these type commute with respect to the sensors. On the contrary, all other degrees of freedom are associated to displacements, and thus will eventually induce non-zero commutation residues when played in different orders from the same starting situation. Hence the results are consistent with the conjecture, and can be used by the agent to autonomously discover which of its actions are associated to displacements or not.

SCP is inversely proportional to how each degree of freedom affects the environment. By that we mean that from the computation of the SCP, we obtain a hierarchical organization of the action space in which the less important dimensions for manipulation and navigation are separated from the dimension that are not crucial for such tasks. For instance, we inferred that shoulders should have a lower SCP than elbows since activating the shoulder joint is more likely to induce non commutativity by moving things around or hitting walls/obstacles. This intuition is verified by our results. Shoulders and base movement have a lower SCP than elbows which in turn have a lower SCP than eyelid and head rotation, as observed in Fig.3. Without having any prior knowledge about the simulation, we can automatically organize the agent's degrees of freedom in a hierarchy. Moreover, the symmetry of the action space is kept, as elbow 1 and 2 have equal SCP, and so do shoulder 1 and 2.

In additional experiments presented in App.C, we verified the robustness of these results. We computed the SCP for 8 different combinations of agents and environments (longer/smaller arms, more/less objects) and confirmed our intuitions on the interpretation of SCP described above.

Alternative methods are not adapted. Details for these two experiments are available in App.B and results are illustrated in Fig.3.

In additional experiments presented in App.D, we performed RL experiments where we have been able to improve sample-efficiency using the SCP computed in this section.

VI. Sensory Commutativity Probability in Realistic Simulators

In this experimental section we compute and interpret the SCP for a realistic embodied agent scenario using the interactive Gibson environment (iGibson) [24].

Figure 4: **Left:** External view of the iGibson simulator where the Fetch robot is in a living room. **Middle:** Fetch's first person view. **Right:** SCP computed for each of Fetch's degrees of freedom.

A. Experimental setup

Simulation description. iGibson is a simulation environment for robotics providing fast visual rendering and physics simulation. It is packed with a dataset with hundreds of large 3D environments reconstructed from real homes and offices, and interactive objects that can be pushed and actuated. In our experiments we use the Rs environment, which is basically a regular apartment. We place the Fetch robot in this environment, see Fig.4. Fetch is originally a 10-DOF real robot [23] equipped with a 7-DOF articulated arm, a base with two wheels and a liftable torso. Fetch perceives the environment through a camera placed in his head, see Fig.4.

SCP computation. In the Flatland environment, two action sequences commuted only if the sensory result of applying both from the same starting situation was perfectly equal. We relax the strict equality condition to compute the SCP for Fetch. Indeed, with real images, only an offset of one pixel would render the two action sequences non sensory commutative. Instead of using the mean squared error as a distance, we use a perceptual distance using the VGG16 [20] features of each observation. We thus have $d(s_1, s_2) = ||VGG16(s_1) - VGG16(s_2)||_2^2$. The choice of the threshold t is arbitrary, we verify in our experiments that a large choice of t leads to equivalent results.

B. Results

The results, presented on Fig.4, are consistent with the Flatland results. Indeed, **the results are consistent with Philipona's conjecture.** The torso lift DOF is not associated with displacement in the environment, so it has a SCP of 1, i.e. it always sensory commutes. Moreover, **SCP is inversely proportional to how each degree of freedom affects the environment.** The wheels have the lowest SCP since they provide longitudinal movement and rotations for the robot. Then comes the first DOF of the articulated arm, i.e. the ones that are closer to its base (like shoulders vs. elbows in the Flatland experiments). Finally the highest SCP values correspond to the arm DOF that are the further on its arm and the torso lift. Once again, we obtain a hierarchical organization of the action space in which the less important dimensions for manipulation and navigation are separated from the dimension that are not crucial for such tasks.

In additional experiments presented in App.E, we verified the robustness of these results by computing the SCP for a different type of robot called JackRabbot [14]. We reach the same conclusions as with the Fetch robot.

VII. Conclusion

We studied the sensory commutativity of action sequences for embodied agent scenarios theoretically. We derived the Sensory Commutativity Probability criterion, which we showed is good proxy for estimating the effect of each action on the environment. We illustrated the potential usefulness of such criterion by improving sample-efficiency in a Reinforcement Learning problem.

References

[1] Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. Exploration by random network distillation. arXiv preprint arXiv:1810.12894, 2018.

[2] Hugo Caselles-Dupré, Louis Annabi, Oksana Hagen, Michael Garcia-Ortiz, and David Filliat. Flatland: a lightweight first-person 2-d environment for reinforcement learning. arXiv preprint arXiv:1809.00510, 2018.

[3] Hugo Caselles-Dupré, Michael Garcia-Ortiz, and David Filliat. Symmetry-based disentangled representation learning requires interaction with environments. In NeurIPS, 2019.

[4] Cédric Colas, Olivier Sigaud, and Pierre-Yves Oudeyer. How many random seeds? statistical power analysis in deep reinforcement learning experiments. arXiv preprint arXiv:1806.08295, 2018.

[5] Dibya Ghosh, Abhishek Gupta, and Sergey Levine. Learning actionable representations with goal-conditioned policies. arXiv preprint arXiv:1811.07819, 2018.

[6] Irina Higgins, David Amos, David Pfau, Sebastien Racaniere, Loic Matthey, Danilo Rezende, and Alexander Lerchner. Towards a definition of disentangled representations. arXiv preprint arXiv:1812.02230, 2018.

[7] Ashley Hill, Antonin Raffin, Maximilian Ernestus, Rene Traore, Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, et al. Stable baselines, 2018.

[8] Donald D Hoffman. The interface theory of perception. Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, 2:1–24, 2018.

[9] Alban Laflaquière. Unsupervised emergence of spatial structure from sensorimotor prediction. arXiv preprint arXiv:1810.01344, 2018.

[10] Alban Laflaquiere, Sylvain Argentieri, Olivia Breysse, Stéphane Genet, and Bruno Gas. A non-linear approach to space dimension perception by a naive agent. In 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 3253–3259. IEEE, 2012.

[11] Alban Laflaquiere, Alexander V Terekhov, Bruno Gas, and J Kevin O'Regan. Learning an internal representation of the end-effector configuration space. In 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 1230–1235. IEEE, 2013.

[12] Alban Laflaquière, J Kevin O'Regan, Sylvain Argentieri, Bruno Gas, and Alexander V Terekhov. Learning agent's spatial configuration from sensorimotor invariants. Robotics and Autonomous Systems, 71:49–59, 2015.

[13] Alban Laflaquière, J Kevin O'Regan, Bruno Gas, and Alexander Terekhov. Discovering space—grounding spatial topology and metric regularity in a naive agent's sensorimotor experience. Neural Networks, 105:371–392, 2018.

[14] Roberto Martín-Martín, Hamid Rezatofighi, Abhijeet Shenoi, Mihir Patel, JunYoung Gwak, Nathan Dass, Alan Federman, Patrick Goebel, and Silvio Savarese. Jrdb: A dataset and benchmark for visual perception for navigation in human environments. arXiv preprint arXiv:1910.11792, 2019.

[15] J Kevin O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. Behavioral and brain sciences, 24(5):939–973, 2001.

[16] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 16–17, 2017.

[17] David Philipona. Développement d'un cadre mathématique pour une théorie sensorimotrice de l'expérience sensorielle. 2008.

[18] Henri Poincaré. L'espace et la géométrie. 1895.

[19] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347, 2017.

[20] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[21] Alexander V Terekhov and J Kevin O'Regan. Space as an invention of active agents. Frontiers in Robotics and AI, 3:4, 2016.

[22] Valentin Thomas, Jules Pondard, Emmanuel Bengio, Marc Sarfati, Philippe Beaudoin, Marie-Jean Meurs, Joelle Pineau, Doina Precup, and Yoshua Bengio. Independently controllable features. arXiv preprint arXiv:1708.01289, 2017.

[23] Melonee Wise, Michael Ferguson, Derek King, Eric Diehr, and David Dymesich. Fetch and freight: Standard platforms for service robot applications. 2016.

[24] Fei Xia, William B Shen, Chengshu Li, Priya Kasimbeg, Micael Edmond Tchapmi, Alexander Toshev, Roberto Martín-Martín, and Silvio Savarese. Interactive gibson benchmark: A benchmark for interactive navigation in cluttered environments. IEEE Robotics and Automation Letters, 5(2):713–720, 2020.

APPENDIX

A. Proofs

Proposition (Structure of $(Seq(\mathcal{M}), \sim, \circ)$)**.**

1) *$\sim$ is an equivalence, i.e. it is reflexive, transitive and symmetric.*
2) *$(Seq(\mathcal{M}), \circ)$ is a group w.r.t $\sim$.*
3) *$\circ$ is not commutative with respect to $\sim$, i.e. we don't generally have $g \circ h \sim h \circ g$.*

Proof:

1) $=$ is an equivalence, thus $\sim$ is an equivalence as well.
2) All 4 properties of the group definition are satisfied. 1. For two action sequences $(h, g) \in Seq(\mathcal{M})$, the composition of h and g is still an action sequence $h \circ g \in Seq(\mathcal{M})$. 2. $\circ$ is associative with respect to $=$, i.e. $g \circ (h \circ k) = (g \circ h) \circ k$ thus it follows that $g \circ (h \circ k) \sim (g \circ h) \circ k$. 3. The identity element is the no-op action. 4. If we suppose that there are no irreversible phenomenons in the environment, then for a fixed (m_t, ϵ_t), all action sequences can be inverted.
3) $\circ$ is not commutative, as we can always explicitly find two action sequences that do not commute. For instance once there exists a movable object in the environment: if the agent is placed left to the object, then let h be moving right and g be moving left. h and g do not commute.

■

B. Alternative methods description

The SCP criterion derived in this paper estimates how much each degree of freedom affects the environment in an embodied agent scenario. In this section we discuss why other approaches cannot reliably estimate the same quantity.

Naive approach: changes in sensors. A straightforward approach to this problem would be to play action sequences of each degree of freedom and quantify how much the sensors change. We consider the squared difference for a transition, i.e. the squared difference for two consecutive observations separated by an action sampled from one dimension of the action space. We report the mean squared difference over 100k transitions, for each degree of freedom.

It is clear in our experiment results, shown in Fig.3, that the approach fails. For instance, rotating the head of the agent changes dramatically what the agent sees, even though this degree of freedom does not affect the environment. It would have made sense if we had considered the top view (fully-observable scenario), since rotating the head does not changes the top view a lot. However in the embodied scenario, this strategy is not viable. For the same reason, approaches based only on the changes in the embodied sensors are bound to fail.

Prediction error approach. A more involved approach would be to use prediction on the sensory change caused by each degree of freedom, a common approach used to improve exploration in RL [1, 16]. The DOF that are harder to predict could be the ones affecting the environment the most, and thus being the most important for manipulation and navigation. We tested this alternative in our experiments, by using a feed-forward neural network to predict the next sensor. The neural network takes a concatenation of the sensor and action at time t and predicts the sensor at time $t+1$. We use the same dataset of transitions as in our experiments with the naive baseline (100k transitions for each degree of freedom, 80k for training and 20k for testing). We trained one model for each degree of freedom, using a neural network with two linear hidden layers with the same number of neurons as the input size. We report the excess prediction error on the held-out test set, i.e. the value of the prediction error minus the minimum prediction error among all 8 degrees of freedom. If the method works, higher excess error prediction should indicate a degree of freedom with more effect on the environment.

The results are shown in Fig.3. It turns out that prediction error is not well correlated with how much a degree of freedom is important for navigation and manipulation. For instance, head rotation, which does not affect the environment, is hard to predict: the agent might not know what's outside his field of view. On the contrary, base longitudinal movement affect the environment a lot and is easier to predict than head rotation.

To conclude, in our experiments we did not find any viable strategy to replace the SCP criterion. SCP is able to easily estimate how important a degree of freedom is for acting and navigating in the environment. The other considered baselines do not manage to organize the action space in the same hierarchical way.

C. Additional experiments on Flatland

In our additional experiments on Flatland, we verify some of the intuitions we built with the main experiments on Flatland. For that, we compute the SCP as described in Sec.V for different combinations of agents and environments. The agents and environments tested are displayed on Fig.5: we use environments with different numbers of objects (from empty to 12 objects), and two agents: one with longer arms than the other.

The results are also displayed on Fig.5. Our intuitions are validated since the more objects are place in the environment, the smaller the value of SCP for DOF that correspond to interacting with these objects. For instance in the empty space almost all DOF have a SCP of 1 since there is nothing to interact with but the walls (that's SCP is not perfectly 1 for base movement annd rotation, shoulder and elbow joints).

Also, we notice that if the arms are longer, the SCP for shoulder and elbow joints is consistently lower for each environment. Indeed, there is more chance to interact with objects if the arms are longer, thus inducing a lower SCP.

D. Sensory Commutativity Probability for efficient RL

We now illustrate how SCP can be used for unsupervised exploration, by using it to improve sample-efficiency in a RL setup. For computational reasons, we experiment with the Flatland simulator only.

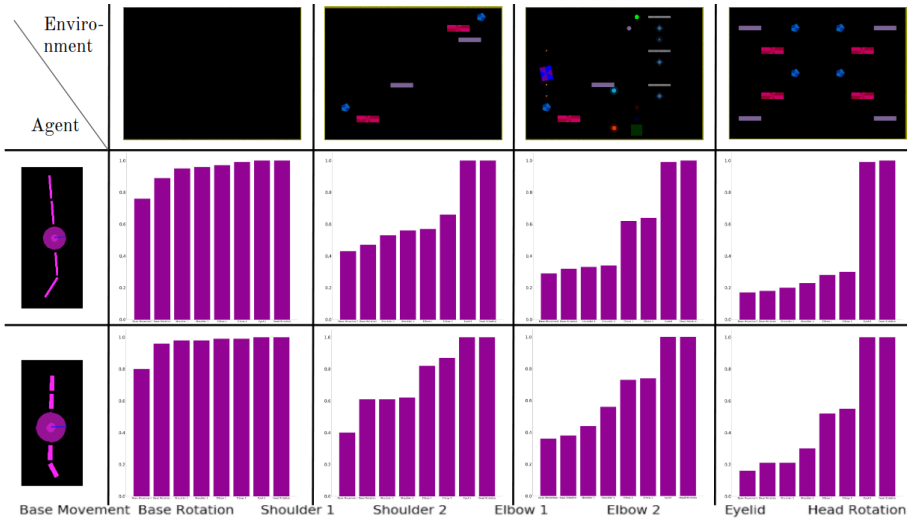

Figure 5: SCP computed for different combinations of agents and environments. Columns: environments. Rows: agents.

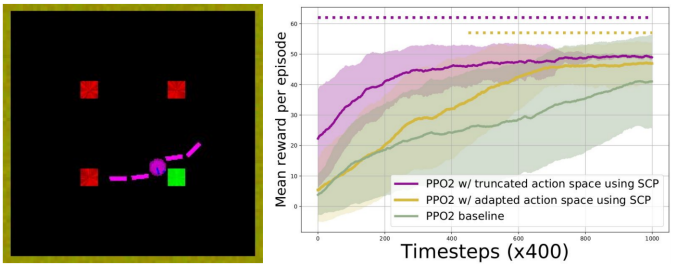

Figure 6: **Left:** RL task. **Right:** Results.

1) Experimental setup: We use the PPO2 [19] implementation from Stable-Baselines [7]. The policy is composed of a 1D convolutional feature extractor followed by a recurrent policy. We consider the same agent, Polyphemus, for which we computed the SCP criterion in Fig.3. The input of the policy is the RGB image of what Polyphemus' eye sees. The environment considered is a square room with 3 dead zones (which terminate the episode with a -20 reward) and a goal zone (which terminates the episode with a +50 reward), illustrated in Fig.2. We propose two methods that take advantage of the SCP to modify the action space of the agent. The goal is to improve sample-efficiency when learning to solve a task in this embodied scenario.

SCP-truncated action space. A first idea is to truncate the agent's action space based on SCP value of each degree of freedom. We implement this by halving the dimension of the action space, keeping only the degrees of freedom that have the most effect on the environment, i.e. lower SCP value. We thus keep the base movement and rotation, and the shoulders joint, while discarding the elbow joints, head rotation and eyelid activation. We refer to this method as *SCP-truncated* action space. This action space reduction will simplify the RL task, as long as the necessary actions such as base motion are selected by the SCP criteria.

SCP-adapted action space. A less involved proposition is to modify the action sampling interval according to the SCP value, for each degree of freedom. This method will modify the exploration dynamics to favor important actions. Suppose that the sampling interval for each dimension of the action space is $[-1, 1]$. If a dimension has high SCP, i.e. it does not affect the environment a lot, we then reduce the interval from which action are sampled $[-1 \cdot l(SCP), 1 \cdot l(SCP)]$. The function l maps the highest SCP to 0 and lowest SCP to 1, then we use a linear interpolation between those two points to deduce values for $SCP \in]-1, 1[$. We refer to this method as *SCP-adapted* action space.

Comparison protocol. We compare those two strategies to a baseline policy trained to solve the task with the complete action space. We average the result of each policy over 30 trials initialized with different random seeds, and we test the statistical significance of our results according to the guidelines provided by [4].

2) Results: The results are displayed on Fig.6. First of all, we notice that all strategies are viable to solve the task. We now compare sample-efficiency between the strategies. The policy trained with *SCP-truncated* action space is able to learn how to solve the task more than twice as fast as the baseline policy. The discarded degrees of freedom are not crucial in this navigation task, hence the agent is still able to solve the task using only the degrees of freedom that have the lowest SCP value. The policy trained with *SCP-adapted* action space is less sample-effective than the *SCP-truncated* but still learns significatively faster than the baseline policy, hence showing our point.

E. Additional experiments on iGibson

We follow the same protocol as with the Fetch robot, i.e. we use the Rs environment and the same algorithm to compute the SCP for the 7 degrees of freedom of the JackRabbot: two wheels and a 5-DOF articulated arm. The results are presented in Fig.7. We observe the hierarchical organization of the DOF of the agent, the wheels having a low SCP as they allow the

robot to move around, and the DOF of the articulated arm having a higher and higher SCP as we move closer to the end of the arm (and thus closer to fine motor skills).

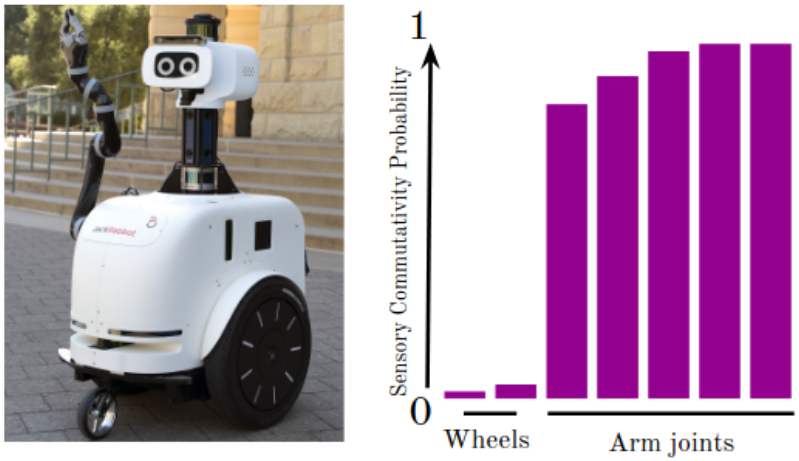

Figure 7: SCP for the JackRabbot (left) in the Rs environment.