SLOW BUT FLEXIBLE OR FAST BUT RIGID? DISCRETE AND CONTINUOUS PROCESSES COMPARED

Matteo Priorelli
Institute of Cognitive Sciences and Technologies
National Research Council of Italy
Padova, Italy

Ivilin Peev Stoianov
Institute of Cognitive Sciences and Technologies
National Research Council of Italy
Padova, Italy
ivilinpeev.stoianov@cnr.it

ABSTRACT

A tradeoff always exists when dealing with multi-step tasks. Higher-level processes can find the best sequence of action primitives to achieve goals in uncertain environments, but they are slow and require significant computational demand. Contrarily, lower-level information processing allows reacting to environmental stimuli rapidly, but the capacity to determine the optimal action when an environmental change occurs is limited. Through reiteration of the same task, biological organisms find the optimal tradeoff: from primitive movements, composite actions gradually emerge by creating low-level task-specific neural structures. What are the underpinnings of such mechanisms and how to exploit them in robotics? We adopted the Active Inference perspective that casts behavior as a minimization of prediction errors, and compared two hierarchical strategies on a pick-and-place task: a discrete-continuous model with global planning capabilities and a continuous-only model with fixed transitions. We show that different phases of motor learning can be expressed in these terms, and propose how discrete actions might be encoded into continuous representations, which helps solve the task fast, smoothly, and with low computational burden. Overall, our study paves the way to understanding task-specialization mechanisms and how they can be adopted in intelligent agents.

1 Introduction

Real-world tasks generally involve complex actions consisting of multiple primitive movements. Even a simple pick-and-place operation consists of opening the hand while moving toward an object, grasping it, moving the arm to a desired position, and finally releasing it. How does the brain support the efficient execution of such complex actions? While a sequence of movements could be easy to realize in static contexts, a difficulty emerges when acting in a dynamic environment, e.g., when a moving object has to be grasped on the fly. Tackling complex multi-step tasks is known to occur deep in the cortical hierarchy, by areas processing discrete, slow-varying entities [1]. But the advantage of discrete representations – highly useful for planning [2] – comes at a cost: if an object moves too fast, deep processing might be too slow and make the grasping action fail.

Motor skill consolidation is a well-known mechanism, most clearly evident in athletes [3, 4, 5]: during repeated exposure to a sequence of actions, an initial learning phase gradually leaves its place to autonomous movements as the athlete becomes more proficient [6]. What seems to happen is that cortical involvement is gradually reduced as specialized neural structures are constructed and take place in lower, fast subcortical regions [7]. The latter cannot extract rich and invariant representations of the environment to perform high-level decision-making but they react rapidly to the sensorium. The basal ganglia are known to be central to task specialization as the striatum can encode chunked representations of action steps that compose a learned habit [8, 9]. But this gradual change seems to be much more pervasive and involve information processing within the cortex itself, as evidenced by a general activity shift from anterior to posterior regions [10]. Nonetheless, the underlying brain mechanisms are still unclear, and considering the frequency of dynamic tasks in real-world scenarios, understanding their computational basis is compelling for robotics.

An interesting proposal is that the brain maintains an internal model of the task dynamics, and motor skill consolidation consists of iteratively fine-tuning this model to achieve a stable behavior that accounts for environmental uncertainty [11, 12]. This hypothesis fits well with a recent brain theory called Active Inference, which brings fundamental insights of increasing appeal into the computational role and principles of the nervous system [13, 14, 15]. Active Inference provides a formalization of the two essential components of control, sensing and acting, both of which are viewed as

aiming to resolve the critical goal of all organisms: to survive in uncertain environments by operating within preferred and expected states (e.g., maintaining a constant temperature). As in Predictive Coding [16], Active Inference assumes that organisms perceive the environment relying on an internal generative model that is constructed by inferring how hidden causes produce sensations [14, 17]. Perception and action are then seen as implementing a dynamic minimization of a quantity called *free energy*, which ultimately translates to minimizing prediction errors: while perception gradually adjusts internal expectations to match the sensory evidence, action gradually samples those sensations that make the expectations true [14, 17].

The frameworks of this theory are suitable for analyzing both facets of motor skill learning under a single perspective. On one hand, Active Inference in continuous time can solve motor control tasks consisting of simple trajectories by defining an internal dynamic model about external (e.g., visual) targets [17]. Elementary movements are then resolved by local suppression of the generated proprioceptive prediction errors [18] which ultimately force a change in the physical state of the actuators. However, when it comes to advanced movements consisting of multiple steps, the continuous framework falls short since the minimization only deals with present sensory signals.

At the other extreme, planning and decision-making are well handled by the Active Inference formulation on discrete state-spaces [19, 20, 21], which views planning as an inferential process [22, 23] and exploits the Partially Observed Markov Decision Process (POMDPs) structure to make predictions about possible combinations of future states, eventually choosing the best sequence of actions for a goal. This is achievable because discrete models can minimize the free energy of expected observations [24, 25, 26, 27], and it seems that operations of cortical areas can be well approximated by such models [1] with higher biological plausibility with respect to state-of-the-art methods such as reinforcement learning [28, 29]. Needless to say, the downside of using discrete representations of the environment is the lack of direct interaction with real-world environments.

Between the two stands a third, less known, architecture: a hybrid framework, which is nothing else but a discrete model used in combination with its continuous counterpart [30, 31]. This involves the transformation of continuous signals into discrete messages, which has been possible through the so-called Bayesian Model Reduction (BMR) [32, 33]. Solving real-world multi-step tasks with a hybrid model consists of chunking the overall action in a series of elementary trajectories, each of which is assigned to a specific discrete state for planning in the future and linearly solvable as a subgoal in the continuous framework. This approach has not received adequate attention yet, as just a few implementations can be currently found in the literature [30, 31, 34], neither of them prospecting real-world applications in dynamic contexts.

Nonetheless, the hybrid model could be key to grasping the virtuous cycle of motor skill learning, which begins with bottom-up messages constructing elementary discrete actions and culminates in top-down signals leading to the formation of task-specific structures. The discrete computations of a hybrid model can simulate the behavior of an agent that has no prior idea of what sequence of actions to take for achieving a goal, that is, when it must compute action "policies" online. However, as soon as the agent has learned a policy, which happens to correspond to a sequence of a-priori defined actions that can take into account some degree of environmental variability, it can adapt by encoding the transitions between discrete states into a continuous dynamics, which then represents a learned advanced motor skill.

For this reason, we used this framework to compare and analyze the behavioral differences in tackling multi-step dynamic tasks, showing that discrete actions and continuous trajectories may work under the same transition mechanisms, which provides a cue for future work on how task specialization could emerge. To this aim, we considered a classic real-world pick-and-place operation. The agent's body is an 8-DoF simulated robotic arm with two fingers, each composed of 2 joints. At each trial, a target object spawns with random and unknown position and dimension, and the agent has to reach the target, close the fingers to simulate the grasping action, reach a randomly specified goal position, and open the fingers to place the target. To verify the capacity of the model to behave correctly in dynamic environments, at each trial a velocity with a random direction and amplitude is assigned to the target, so that the grasping action might fail if the target moves too fast.

2 Results

2.1 The phases of motor learning

It has been hypothesized that motor skill learning is roughly composed of three different phases [35]: a first cognitive phase where movements are inefficient and one has to consciously explore different strategies to achieve a goal; an associative stage where the learner becomes more proficient and transitions between movements are more fluid; a final autonomous phase where the learner has reached almost complete autonomy from conscious task-specific thinking and can move without cognitive effort. The Active Inference theory provides interesting insights about this process because: (i) it builds upon an internal generative model of the task dynamics, which can be constantly refined by experience;

(ii) it assumes that the whole information processing starts and unfolds through the same principle of prediction error minimization, either with continuous or discrete representations.

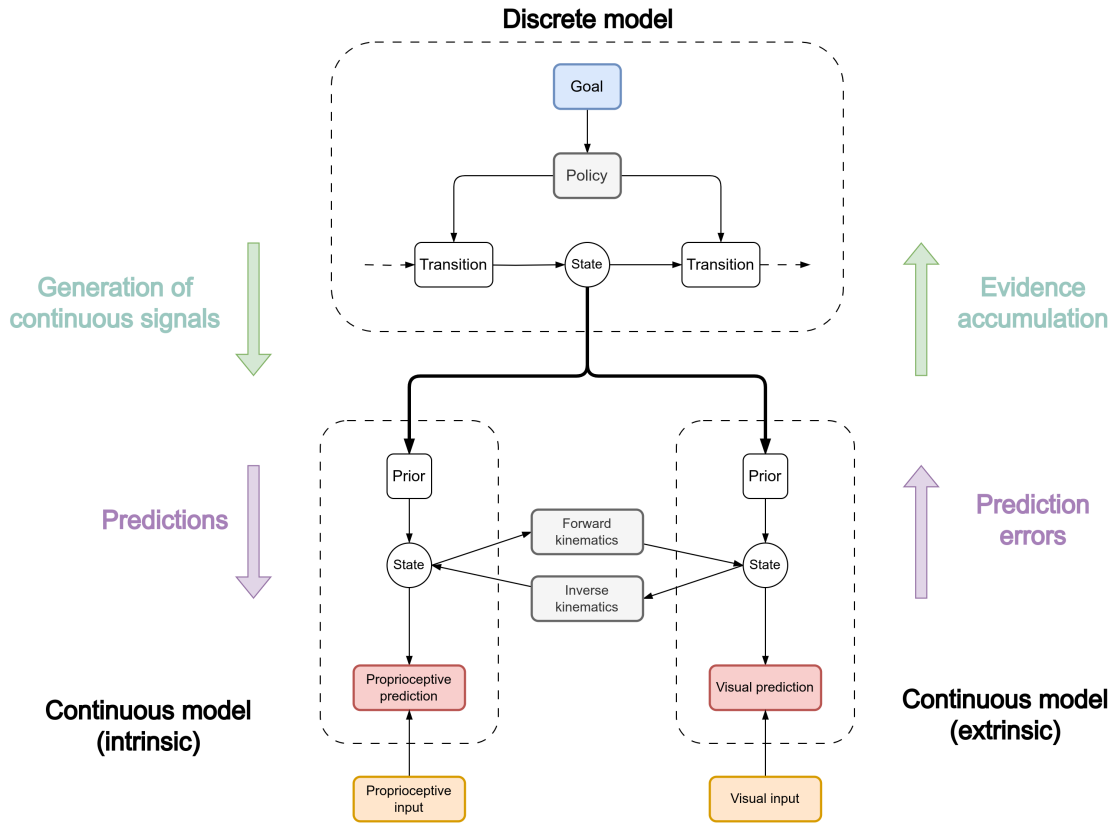


Figure 1: Graphical representation of the first cognitive phase. In the first trials of every novel task, the state transitions generally encode elementary actions, such as reaching, grasping, and releasing. The cognitive effort needed by high-level processes to learn the correct sequence of elementary actions translates to the task dynamics being only visible in the discrete model, while the continuous levels only receive a prior representation and their job is mostly to realize the corresponding trajectories. This is in line with the hypothesis that in the first phase of motor learning a high cortical activity is recorded, especially in anterior (e.g., prefrontal) regions.

Under this perspective, neural processing occurring in the first cognitive phase can be well captured by the hybrid architecture shown in Figure 1, composed of a discrete and two continuous models. The role of the discrete model is to find the optimal *policy* or sequence of actions to reach a *goal* based on a discretized representation of the environment dynamics. Since there are several differences with respect to conventional theories of motor control, we provide an overview of the key concepts regarding high-level and low-level processes. For a detailed technical description, we remind to the Methods section.

First, the goal is encoded as a prior belief over the policy that the agent has to learn: since the policy shapes the discrete dynamics encoding how the world evolves, this means that if one has the false belief to be in a particular state, his or her actions will eventually result in looking for those states that make his or her beliefs true.

Second, each level is constantly involved in computing a prediction, comparing it with the state below, and updating its state with the generated prediction error. These simple steps are repeated throughout the whole hierarchy so that, from a discrete goal, a cascade of primitive proprioceptive trajectories is generated that is eventually realized by the lowest levels of the motor system.

Third, these primitive trajectories are generated by a continuous model encoding *intrinsic* kinematic (e.g., joint angles) information; however, the movements that the agent wants to realize are usually thought of in an *extrinsic* (e.g., visual) domain. As in optimal control, these two are related by *forward kinematics*, but here the intrinsic representation is found by inferring the most likely kinematic configuration that may have generated the current extrinsic state [36]. These two continuous modalities are needed even in a pick-and-place task because reaching movements usually occur

in an extrinsic reference frame, while opening and closing the hand are operations better understood from an intrinsic perspective.

Fourth, top-down communication between discrete and continuous frameworks is made by associating each discrete state to some continuous prior, and by averaging between all discrete probabilities. Similarly, bottom-up communication is possible through the comparison of the continuous state with each continuous prior to find the most likely discrete representation that may have generated the current trajectory.

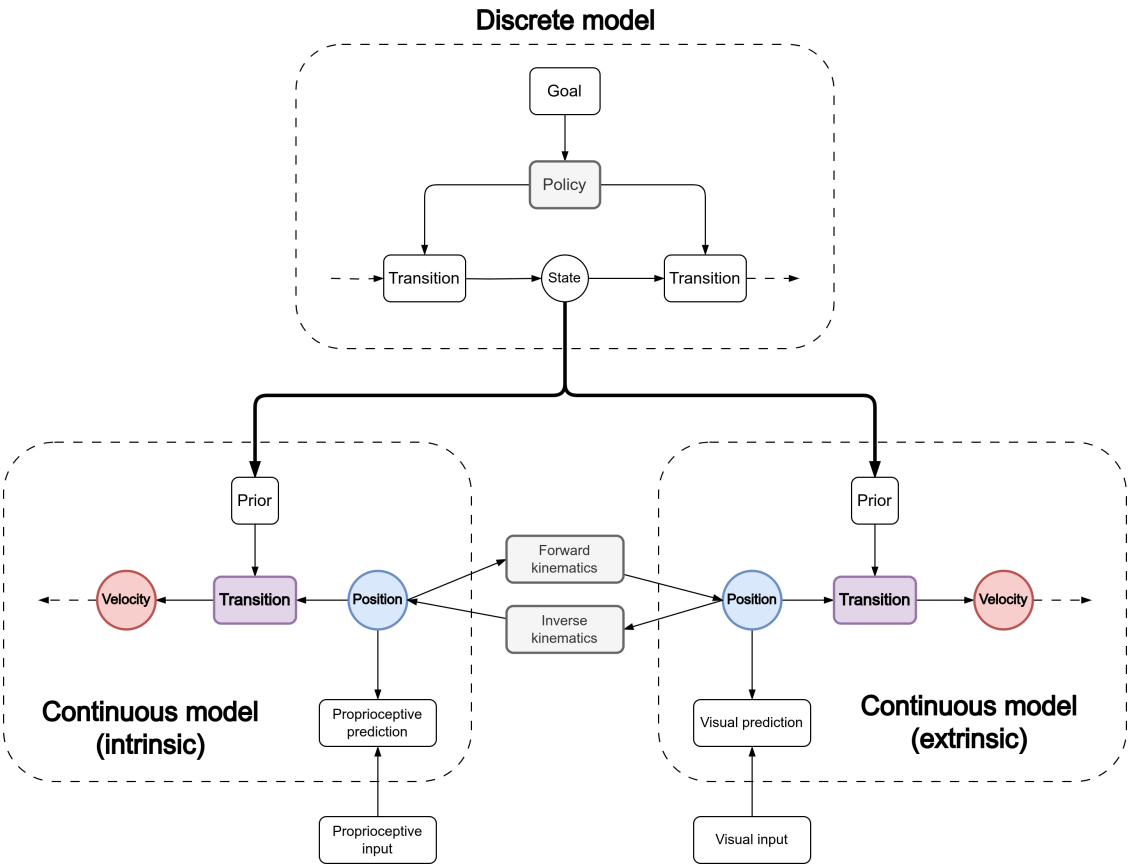


Figure 2: Graphical representation of the second associative phase. A continuous dynamics is represented in generalized coordinates of motion including position, velocity, acceleration, etc., allowing one to express with increasing fidelity the true environmental dynamics. Further, the outcomes of the discrete model not only biases the 0th temporal order (i.e., position) state, but steer the whole continuous trajectory toward a particular direction. In our reach-to-grasp task, this might correspond to considering the elementary actions of reaching and grasping as single discrete states that generate composite trajectories in the continuous domain.

In the next, associative phase, the cortical activity begins to shift toward posterior areas. The process can be modeled as in Figure 2, with a more balanced work between discrete and continuous models in which the latter start to construct their own task dynamics, while in the previous case they only received a static prior resulting from a combination of all discrete states.

This intermediate stage provides insights into both the transition mechanism and how the dynamics is encoded. While in the cognitive phase the task would be comprised of a reaching movement, a cognitive effort to find the next optimal action, and then a pure grasping action, in the associative phase the continuous models would handle the transition generating a fluid movement that makes the hand close as soon as it approaches the target. The discrete model can hence represent the task with a reaching-grasping action and a reaching-releasing action, with reduced computational demand.

The increased efficiency is also possible because in Figure 1 the discrete states had to comprise every possible elementary movement – either intrinsic or extrinsic. In this case, each modality begins to encode its own dynamics, translating to grasping (or reaching) movements being mapped mostly into the intrinsic (or extrinsic) modality.

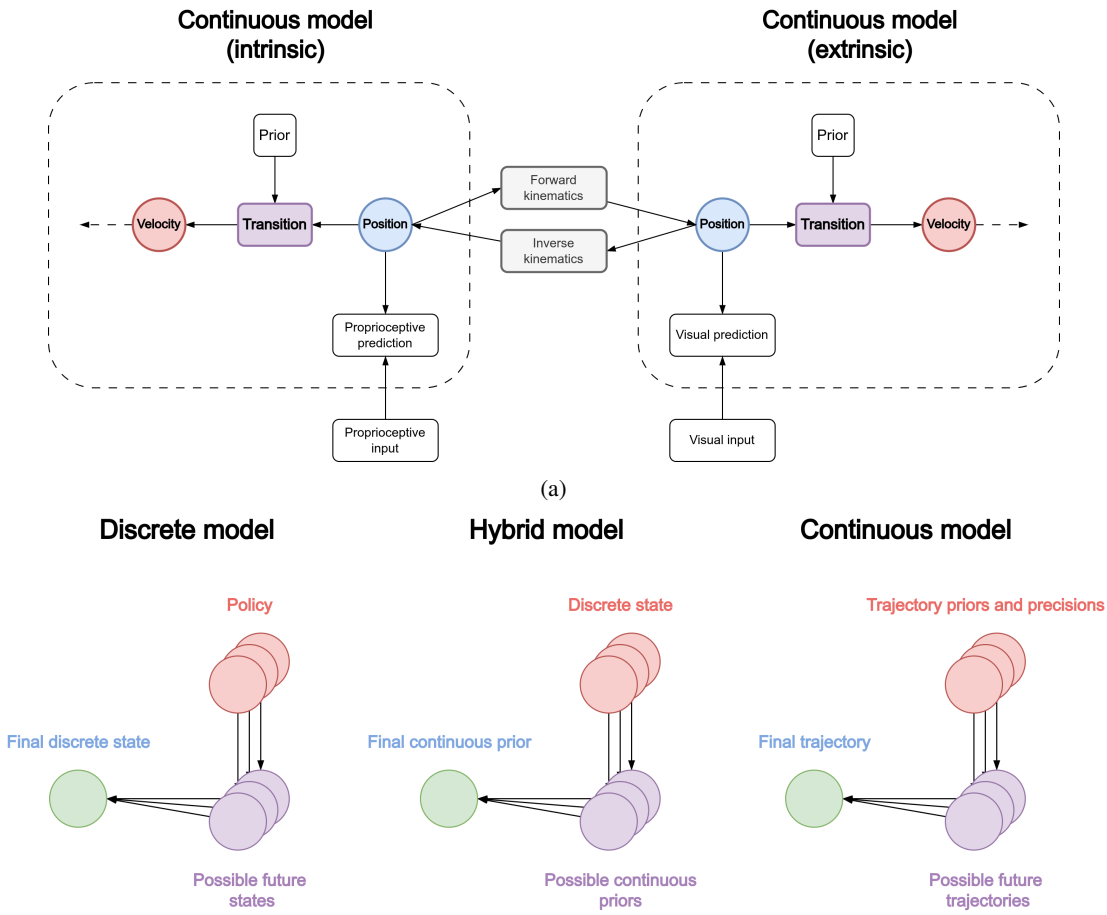


Figure 3: (A) Graphical representation of the third autonomous phase. The process of task specialization is complete. The discrete model – not shown here – has learned a representation abstract enough to include all the four main movements of the pick-and-place operation and its role is only to command its execution and accumulate evidence from both modalities to infer when it has been completed. (B) Comparison between discrete, hybrid, and continuous models.

Figure 3a finally shows the third, autonomous phase of motor learning. The continuous models now encode the dynamics of the entire task in the most efficient way because the mechanism is fully autonomous and does not require the repeated activation of higher levels for planning the next action.

But how are transitions between continuous trajectories encoded in practice? To elucidate, we reveal an interesting parallelism between high-level and low-level processes: in the discrete model, states are conditioned on every action that the agent can take in a specific instant through corresponding probability distributions, and the overall state is found by averaging the expected states from all transitions. The same technique is applied to hybrid models, where the probability of every possible discrete outcome is combined with a continuous state to compute the average signal that will bias the lower levels. Similarly, we propose that a continuous dynamics can be generated by averaging the results from independent probability distributions, each encoding a particular trajectory.

The consequence is that goal-directed behavior is achieved through specular mechanisms at different hierarchical locations, as shown in Figure 3b: however, in the continuous-only model the future trajectories are weighted by two different processes. A *fast* process that imposes and infers the actions of the discrete model (shown in Figure 2); but also a *slow* process that learns the *precisions* – or inverse variances – of every trajectory. As explained in the Methods section, this is a direct consequence of how the prediction error of each trajectory drives the update of the latent states. This result lends itself to an intuitive interpretation: in Active Inference, a low precision in a specific sensory modality implies that it cannot be trusted and the agent should update its internal state relying on other signals. Correspondingly,

a trajectory with a high precision will be a good option for minimizing the prediction errors in the current context, that is, the agent will be confident in using it for the task considered.

2.2 Discrete and continuous processes compared

To analyze the behavior under the different stages of motor learning, we used the same four elementary actions and adopted the following method: during the autonomous phase, the agent was controlled by a continuous-only model with fixed transitions among trajectories, where their priors depended on task-specific conditions over the continuous latent states. In other words, the agent used an a-priori (assumed to be learned) sequence of actions and it could not replan if unexpected conditions occurred; however, the strength of each trajectory could vary, and it is indeed this semi-flexible dynamics that allow performing smoothly composed primitive movements in highly dynamic contexts. As concerns the other two learning phases, we used a hybrid model and compared the performances with different sampling time windows for each discrete operation at the upper level. Specifically, this time window is needed to accumulate evidence from the lower-level continuous modalities and plan the next primitive movement.

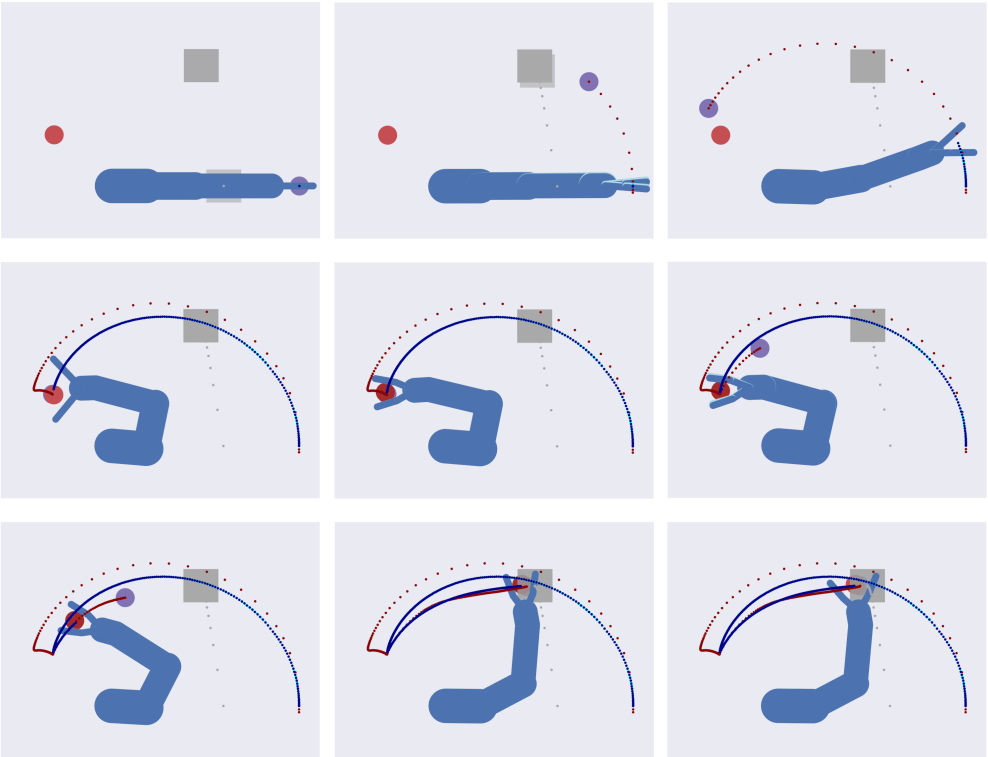
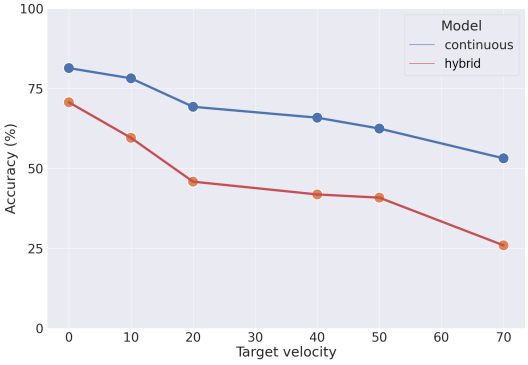


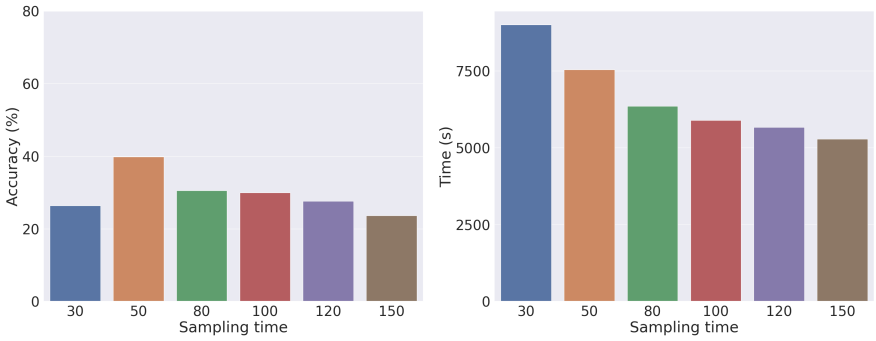
Figure 4: Sequence of time frames of the grasping task. Real and estimated target positions are represented by red and purple circles, real and estimated goal positions by grey and dark grey squares, and real and estimated limb positions in blue and cyan. Hand, target, and goal belief trajectories are represented respectively with a blue, red, and grey dotted line.

A sequence of time frames is shown in Figure 4 for a sample grasping trial with continuous-only control. Note that both the intrinsic and extrinsic latent states (or *beliefs*) comprise estimates of the arm, the target to be grasped, and the goal position, as described in the Methods section.

Figure 5a represents the capacity of the continuous and hybrid models to solve the task in dynamic conditions, using a fixed sampling time window for the hybrid model. While the two models achieve somewhat similar performances in static grasping (i.e., zero target velocity), the difference increases when it comes to grasping a moving target. The performance drop of the hybrid model is caused by the amount of sampling of continuous evidence performed by the discrete part, as illustrated in Figure 5b, which reveals a tradeoff between reaction times and model complexity. Finally, Figure 6 shows the interplay between reaching and gasping actions with the hybrid model and a constant sampling time, where five different phases emerge from the four discrete actions.



(a)

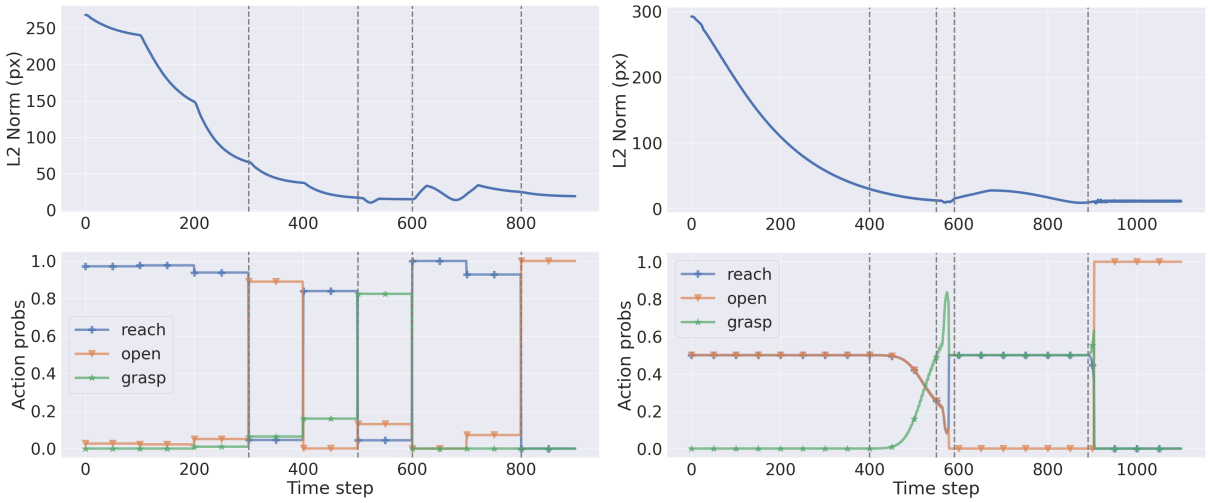


(b)

Figure 5: (A) Performance comparison between the hybrid and continuous-only models for different target velocities. Accuracy measures the number of successful trials. As the velocity of the target increases, the grasping action becomes more difficult for the hybrid model since faster response planning is needed. Instead, the continuous-only agent constantly adjusts its priors based on the continuous evidence inferred at each time step, and therefore it handles dynamic tasks more smoothly and efficiently. (B) Behavioral effects of the sampling time in the hybrid model with a target moving at a moderate velocity. Accuracy (left) and overall simulation time (right) are shown. Every condition is averaged over 1000 trials. On one hand, greater sampling time allows the agent to accumulate more information and hence estimate more accurately the discrete state. However, longer sampling time also implies chunking the action into a smaller number of longer primitive actions, which ultimately makes the grasping action fail since the moving target is continuously changing its location during the sampling period. On the contrary, reducing the sampling time has the advantage of speeding up the agent's planning process (i.e., lower reaction times) so that it can grasp the target with greater accuracy, but only up to a point. In this case, maintaining the same performances would require longer policies due to a greater number of continuous trajectories needed for a reaching movement. At the same time, since the discrete model is activated at a higher rate for action replanning, the simulation time steadily increases as well, which can be associated with higher energy demands.

3 Discussion

Understanding the mechanisms that support task specialization and motor skill learning is compelling for making advances in current robots and intelligent systems. However, a common issue is that high-level and low-level processes are often analyzed and developed with different perspectives and techniques, e.g., optimal control algorithms [37] or deep neural networks [38]. Instead, the flexibility and robustness behind human and animal motor systems reside in the fact that the learning of a new skill is a unified process that initially involves significant activation of prefrontal brain areas gradually shifting toward posterior and subcortical regions [10]. Indeed, constantly relying on the whole hierarchy is computationally demanding and high-level areas are unable to adapt to environmental changes rapidly. Highly dynamic tasks can be therefore solved more efficiently by pushing the learned action transitions and policies down to lower, faster hierarchical layers operating in continuous domains [39, 10]. The Posterior Parietal Cortex is known to encode multiple goals in parallel during sequences of actions as in multi-step reaching, even when there is a considerable delay between goal states [40], and evidence suggests that rhythmic and repetitive movements do not involve the continuous activation of prefrontal areas but only rely on sensorimotor circuits [41].



(a) Action probabilities of the hybrid model.

(b) Trajectory priors of the continuous model.

Figure 6: Distance between hand and target (top), and strengths of reaching and grasping actions (bottom), for both the hybrid (left panel) and the continuous (right panel) models. Notably, five different phases can be distinguished: a pure reaching movement, an intermediate phase when the agent slowly approaches the target and prepares the grasping action, a grasping phase, another reaching movement and, finally, the target release. The computation of the action probabilities by the discrete model thus realizes a smooth transition between reaching and grasping actions, although at this stage still encoded as separate representations. This might provide a cue about how composite movements naturally emerge in lower levels during intermediate phases of motor skill learning. Similar behavior can be easily realized by the continuous model if the trajectory priors depend on sigmoid functions of the beliefs.

We propose that transition mechanisms in motor skill learning could be explained from a unified Active Inference perspective. When the agent is required to interact with a novel situation, high-level planning is essential because the low levels are not able to minimize the generated prediction errors, which then climb up the cortical hierarchy. As the agent becomes more confident and learns composite state transitions that already take into account environmental uncertainty and possible dynamic elements, the prediction errors arising during the unfolding of the task can be explained away more efficiently by lower-level specialized neurons, and repeated calling of high levels ceases. Even when using elementary discrete actions, a composite movement corresponding to an approaching phase between reaching and grasping naturally arises from the Bayesian Model Reduction technique because the continuous evidence accumulation provides a smooth transition between the probabilities over the discrete hidden states, as shown in Figure 6. We therefore propose that the dynamics of a continuous model may have a specular structure to its discrete counterpart – i.e., the final trajectory is generated by weighting independent distributions. As a result, the composite movement might be embedded into the continuous dynamics with two parallel processes: while a discrete model can rapidly impose and infer particular trajectories, their precisions are additionally adapted through the same mechanisms of free energy minimization, so that the system will score how well an action can perform. In other words, we propose that these precisions act not just as modulatory signals but represent the confidence of the agent [42] about the current state of the task, in a specular manner to sensory precisions. The higher the precision of a specific trajectory, the more useful it is in explaining the agent's final goal; on the other hand, a low trajectory precision means that the agent is not confident about its selection for a particular context.

Further, the results in Figure 5b show an interesting tradeoff between reaction times, planning capabilities, and computational demand. If the environment changes with high frequency, occasional calling of the discrete model does not permit to respond in time to novel sensory observations. On the other hand, repeated action replanning is counterproductive beyond a certain limit since, although allowing the agent to react more rapidly to environmental changes, it increases the effort of the discrete model to find the correct policy, as a higher number of discrete steps are needed. Instead, a continuous-only model achieves the best performances in less time and with small resource expenditure, but at the inability to perform replanning if the task goes into unexpected tracks.

As shown in the Methods section, the presented work also contains two main novelties with respect to state-of-the-art hybrid models: first, the reduced priors are generated dynamically at the beginning of each discrete step through the specification of flexible intentions – which roughly correspond to the discrete actions – letting the agent able to operate in dynamic environments. Second, the discrete model can impose priors and accumulate evidence over

multiple continuous modalities, e.g., intrinsic and extrinsic coordinates. This has the advantage that one can realize more complex goals (e.g., reaching with specific affordances, as shown in the Supplementary Materials) and has access to more information for inferring the combined discrete state (e.g., extrinsic coordinates for the position, and intrinsic coordinates for the hand status).

Last, the analogies that we presented between discrete, mixed, and continuous models may shed light on the information processing within cortical regions. It has been hypothesized that the cortex works within a system of discrete models of increasing hierarchical depth (e.g., motor, associative, limbic) [9], while the interface with the continuous models that receive sensory observations is supposed to be achieved through subcortical structures, such as the superior colliculus [31] or the thalamus [30]. Since deeper cortical layers are found to encode increasingly discretized representations [15], the latter could also be the consequence of the invariability of neural activity that results during inference when the hierarchy is temporally deep. In this view, whether it makes more sense to consider a specific layer as encoding a discrete or continuous representation – cortical computations have been also simulated by continuous models [42, 43], but as well noted in [15], even the first computations of neural processing could be viewed as dealing with lots of little categories – from a high-level perspective a gradual transition between the two modalities could occur. Therefore, since all regions eventually have to deal with continuous signals, looking for analogies between the two processes might be helpful (e.g., reward-related activity not involving the standard reward pathways has been recorded also in the Primary Visual Cortex V1 [44]).

Overall, these results pave the way for how task-specialization mechanisms could occur in biologically plausible ways. However, the three stages of motor learning were analyzed separately to show the relationships between policies and dynamic trajectories. Further studies are needed to understand how the adaptation from discrete to continuous processes may arise. Two promising directions of research would be to implement deep hierarchical models [36], and to let priors and precisions of the continuous dynamics – which in the simulations were kept fixed throughout the task – adapt to the prediction errors that they receive, thus allowing self-modeling based on the current task and environmental state.

4 Methods

4.1 Perception, Control, and the Variational Free Energy

Active Inference assumes that organisms perceive the environment relying on an internal generative model that is constructed by inferring how external causes produce sensations in the real generative process [14, 17]. In the continuous domain, this process is usually factorized into probability distributions over hidden causes $\tilde{\boldsymbol{v}}$, hidden states $\tilde{\boldsymbol{x}}$ and observable outcomes $\tilde{\boldsymbol{y}}$:

$$p(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{v}}, \tilde{\boldsymbol{y}}) = p(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}})p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{v}})p(\tilde{\boldsymbol{v}}) \tag{1}$$

These distributions are approximated by Gaussian functions:

$$\begin{aligned} p(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}) &= \mathcal{N}(\tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}}), \tilde{\boldsymbol{\Pi}}_y^{-1}) \\ p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{v}}) &= \mathcal{N}(\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{v}}), \tilde{\boldsymbol{\Pi}}_x^{-1}) \\ p(\tilde{\boldsymbol{v}}|\boldsymbol{\eta}) &= \mathcal{N}(\boldsymbol{\eta}, \tilde{\boldsymbol{\Pi}}_v^{-1}) \end{aligned} \tag{2}$$

corresponding to the following non-linear stochastic equations that represent how the environment evolves:

$$\begin{aligned} \tilde{\boldsymbol{y}} &= \tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}}) + \boldsymbol{w}_y \\ \mathcal{D}\tilde{\boldsymbol{x}} &= \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{v}}) + \boldsymbol{w}_x \end{aligned} \tag{3}$$

where $\mathcal{D}$ is the differential shift operator, and the symbol $\sim$ indicates variables encoded in generalized coordinates representing their instantaneous trajectories.

Directly evaluating the posterior $p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$ is intractable since it involves the computation of the inaccessible marginal $p(\tilde{\boldsymbol{y}})$. The variational solution is to approximate the posteriors with tractable distributions, e.g., Gaussians:

$$\begin{aligned} q(\tilde{\boldsymbol{x}}) &= \mathcal{N}(\tilde{\boldsymbol{x}}; \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{P}}_x^{-1}) \\ q(\tilde{\boldsymbol{v}}) &= \mathcal{N}(\tilde{\boldsymbol{v}}; \tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{P}}_v^{-1}) \end{aligned} \tag{4}$$

where the parameters $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\nu}}$ are called beliefs over hidden states and hidden causes, respectively. Perceptual inference then turns into a minimization of the difference between the approximate and real posteriors, which can be formalized in terms of a KL divergence – equivalent to the expectation, over the approximate posterior, of the difference between the two log-probabilities:

$$D_{KL}[q(\tilde{\boldsymbol{x}})||p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})] = \sum\nolimits_x q(\tilde{\boldsymbol{x}}) \ln \frac{q(\tilde{\boldsymbol{x}})}{p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})} = \mathop{\mathbb{E}}_{q(\tilde{\boldsymbol{x}})}[\ln q(\tilde{\boldsymbol{x}}) - \ln p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})] \tag{5}$$

Given that the denominator $p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$ still depends on the marginal $p(\tilde{\boldsymbol{y}})$, we express the KL divergence in terms of the log-evidence and the *Variational Free Energy* (VFE), and minimize the latter quantity instead. Given that the KL divergence is non-negative (due to Jensen's inequality), the VFE provides an upper bound on the log-evidence – thus, VFE minimization improves both evidence and model fit:

$$\mathcal{F} = \mathop{\mathbb{E}}_{q(\tilde{\boldsymbol{x}})}\left[\ln \frac{q(\tilde{\boldsymbol{x}})}{p(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})}\right] = \mathop{\mathbb{E}}_{q(\tilde{\boldsymbol{x}})}\left[\ln \frac{q(\tilde{\boldsymbol{x}})}{p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})}\right] - \ln p(\tilde{\boldsymbol{y}}) \tag{6}$$

Under the previous approximations, the free energy minimization turns into an iterative parameter update through the following expressions:

$$\begin{aligned} \dot{\tilde{\boldsymbol{\mu}}} - \mathcal{D}\tilde{\boldsymbol{\mu}} &= -\partial_{\boldsymbol{\mu}}\mathcal{F} = \partial\tilde{\boldsymbol{g}}^T\tilde{\boldsymbol{\Pi}}_y\tilde{\boldsymbol{\varepsilon}}_y + \partial_{\boldsymbol{\mu}}\tilde{\boldsymbol{f}}^T\tilde{\boldsymbol{\Pi}}_x\tilde{\boldsymbol{\varepsilon}}_x - \mathcal{D}^T\tilde{\boldsymbol{\Pi}}_x\tilde{\boldsymbol{\varepsilon}}_x \\ \dot{\tilde{\boldsymbol{\nu}}} - \mathcal{D}\tilde{\boldsymbol{\nu}} &= -\partial_{\boldsymbol{\nu}}\mathcal{F} = \partial_{\boldsymbol{\nu}}\tilde{\boldsymbol{f}}^T\tilde{\boldsymbol{\Pi}}_x\tilde{\boldsymbol{\varepsilon}}_x - \tilde{\boldsymbol{\Pi}}_v\tilde{\boldsymbol{\varepsilon}}_v \end{aligned} \tag{7}$$

where $\tilde{\boldsymbol{\varepsilon}}_s$, $\tilde{\boldsymbol{\varepsilon}}_x$ and $\tilde{\boldsymbol{\varepsilon}}_v$ are respectively prediction errors of sensations, dynamics and prior:

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}}_y &= \tilde{\boldsymbol{y}} - \tilde{\boldsymbol{g}}(\tilde{\boldsymbol{\mu}}) \\ \tilde{\boldsymbol{\varepsilon}}_x &= \mathcal{D}\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}}) \\ \tilde{\boldsymbol{\varepsilon}}_v &= \tilde{\boldsymbol{\nu}} - \boldsymbol{\eta} \end{aligned} \tag{8}$$

Notably, the VFE can also be used to compute motor control signals:

$$\dot{\boldsymbol{a}} = -\partial_{\boldsymbol{a}}\mathcal{F} = -\partial_{\boldsymbol{a}}\tilde{\boldsymbol{g}}^T\tilde{\boldsymbol{\Pi}}_y\tilde{\boldsymbol{\varepsilon}}_y \tag{9}$$

thus sampling those sensations that correspond to the current agent's representation of the world, i.e., Active Inference. This process further reduces the VFE, allows implementing goal-directed behavior and keeping the agent in predictable and safer spaces [14].

4.2 Planning with the Expected Free Energy

Although Active Inference in continuous time can deal with real-world problems by keeping track of the instantaneous trajectories that the continuous environment produces, it has several limitations and a narrow use as it cannot easily handle more general types of actions, including decision-making. The minimization of the VFE can only adjust the approximate posterior depending on current – or past – observations, and it does not evaluate future states and outcomes. To endow an agent with this ability, a quantity called *Expected Free Energy* (EFE) is considered [19, 20]. We first augment the generative model with policies $\boldsymbol{\pi}$:

$$p(\boldsymbol{s}, \boldsymbol{o}, \boldsymbol{\pi}) = p(\boldsymbol{o}|\boldsymbol{s}, \boldsymbol{\pi})p(\boldsymbol{s}|\boldsymbol{\pi})p(\boldsymbol{\pi}) \tag{10}$$

where $\boldsymbol{s}$ and $\boldsymbol{o}$ are discrete states and outcomes. Note that the policies are not simple stimulus-response mappings – as in Reinforcement Learning schemes – but sequences of actions. Since action planning involves selecting those policies that lead to the desired priors, we also have to consider future outcomes that have not been observed yet. Thus, the EFE is specifically constructed by conditioning over them, which are therefore treated as hidden states:

$$\mathcal{G}_\pi = \mathop{\mathbb{E}}_{q(\boldsymbol{s},\boldsymbol{o}|\boldsymbol{\pi})}\left[\ln \frac{q(\boldsymbol{s}|\boldsymbol{\pi})}{p(\boldsymbol{s}, \boldsymbol{o}|\boldsymbol{\pi})}\right] \approx \mathop{\mathbb{E}}_{q(\boldsymbol{s},\boldsymbol{o}|\boldsymbol{\pi})}\left[\ln \frac{q(\boldsymbol{s})}{q(\boldsymbol{s}|\boldsymbol{o}, \boldsymbol{\pi})}\right] - \mathop{\mathbb{E}}_{q(\boldsymbol{o}|\boldsymbol{\pi})}[\ln p(\boldsymbol{o}|\boldsymbol{C})] \tag{11}$$

where the probability distribution $p(\boldsymbol{o}|\boldsymbol{C})$ encodes preferred outcomes. The last two terms are respectively called *epistemic* (uncertainty reducing) and *pragmatic* (goal seeking). In practice, this quantity is used by first factorizing the agent's generative model as in POMDPs:

$$p(\boldsymbol{s}_{1:T}, \boldsymbol{o}_{1:T}, \boldsymbol{\pi}) = p(\boldsymbol{s}_1) \cdot p(\boldsymbol{\pi}) \cdot \prod_{\tau=1}^{T} p(\boldsymbol{o}_\tau|\boldsymbol{s}_\tau) \cdot \prod_{\tau=2}^{T} p(\boldsymbol{s}_\tau|\boldsymbol{s}_{\tau-1}, \boldsymbol{\pi}) \tag{12}$$

Each of these elements can be represented with categorical distributions:

$$\begin{aligned} p(\boldsymbol{s}_1) &= Cat(\boldsymbol{D}) & p(\boldsymbol{o}_\tau|\boldsymbol{s}_\tau) &= Cat(\boldsymbol{A}) \\ p(\boldsymbol{\pi}) &= Cat(\boldsymbol{E}) & p(\boldsymbol{s}_\tau|\boldsymbol{s}_{\tau-1}, \boldsymbol{\pi}) &= Cat(\boldsymbol{B}_{\pi,\tau}) \end{aligned} \tag{13}$$

where $\boldsymbol{D}$ encodes beliefs about the initial state, $\boldsymbol{E}$ encodes the prior over policies, $\boldsymbol{A}$ is the likelihood matrix and $\boldsymbol{B}_{\pi,\tau}$ is the transition matrix. If we assume, under the mean-field approximation, that the approximate posterior factorizes into independent distributions:

$$\begin{aligned} p(\boldsymbol{s}_{1:T}|\boldsymbol{o}_{1:T}, \boldsymbol{\pi}) \approx q(\boldsymbol{s}_{1:T}, \boldsymbol{\pi}) &= q(\boldsymbol{\pi}) \prod_{\tau}^{T} q(\boldsymbol{s}_\tau|\pi) \\ q(\boldsymbol{\pi}) &= Cat(\boldsymbol{\pi}) \\ q(\boldsymbol{s}_\tau|\boldsymbol{\pi}) &= Cat(\boldsymbol{s}_{\pi,\tau}) \end{aligned} \tag{14}$$

we can use the standard technique of variational message passing to infer each posterior $q(\boldsymbol{s}_\tau|\boldsymbol{\pi})$ and combine them into a global posterior $q(\boldsymbol{s}_{1:T}|\boldsymbol{\pi})$. In order to update the posterior about hidden states, we combine the messages from past states, future states, and outcomes, express each term with its sufficient statistics, and finally apply a softmax function to get a proper probability distribution:

$$\boldsymbol{s}_{\pi,\tau} = \sigma(\ln \boldsymbol{B}_{\pi,\tau-1}\boldsymbol{s}_{\pi,\tau-1} + \ln \boldsymbol{B}^T_{\pi,\tau}\boldsymbol{s}_{\pi,\tau+1} + \ln \boldsymbol{A}^T\boldsymbol{o}_\tau) \tag{15}$$

Similarly, in order to update $q(\boldsymbol{\pi})$, we combine the messages from the prior over policies given by the matrix $\boldsymbol{E}$, and from future observations conditioned upon policies; we can approximate the latter by the EFE conditioned at a particular time τ:

$$\begin{aligned}\mathcal{G}_\pi &\approx \sum_\tau D_{KL}[q(\boldsymbol{o}_\tau|\boldsymbol{\pi})||p(\boldsymbol{o}_\tau|\boldsymbol{C})] + \mathop{\mathbb{E}}_{q(\boldsymbol{s}_\tau|\boldsymbol{s}_{\tau-1},\boldsymbol{\pi})}[H[p(\boldsymbol{o}_\tau|\boldsymbol{s}_\tau)]] \\ &= \sum_\tau \boldsymbol{o}_{\pi,\tau}(\ln \boldsymbol{o}_{\pi,\tau} - \boldsymbol{C}_\tau) + \boldsymbol{s}_{\pi,\tau}\boldsymbol{H}\end{aligned} \tag{16}$$

where:

$$\boldsymbol{o}_{\pi,\tau} = \boldsymbol{A}\boldsymbol{s}_{\pi,\tau} \qquad \boldsymbol{C}_\tau = \ln p(\boldsymbol{o}_\tau|\boldsymbol{C}) \qquad \boldsymbol{H} = -diag(\boldsymbol{A}^T \ln \boldsymbol{A}) \tag{17}$$

Finally, we select the action that is the most likely under all policies:

$$\begin{aligned}\boldsymbol{\pi} &= \sigma(\ln \boldsymbol{E} - \boldsymbol{\mathcal{G}}) \\ u_t &= \arg\max_u \boldsymbol{\pi} \cdot [U_{\pi,t} = u]\end{aligned} \tag{18}$$

4.3 Bayesian Model Reduction in Hybrid Models

To get a discrete model working with the richness of the continuous input, a form of communication with the continuous framework is required. In particular, since the two levels here operate in different domains, we need a way to obtain a continuous prior from a discrete prediction and, concurrently, to estimate a discrete outcome based on continuous evidence. Both problems can be easily addressed through Bayesian Model Reduction [33, 32]. Consider a generative model $p(\boldsymbol{\theta}, \boldsymbol{y})$ with parameters $\boldsymbol{\theta}$ and data $\boldsymbol{y}$:

$$p(\boldsymbol{\theta}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{19}$$

and an additional distribution $\tilde{p}(\boldsymbol{y}, \boldsymbol{\theta})$, which is a reduced version of the first model if the likelihood of some data is the same under both models and the only difference rests upon the specification of the priors $\tilde{p}(\boldsymbol{\theta})$. We can express the posterior of the reduced model as:

$$\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) = p(\boldsymbol{\theta}|\boldsymbol{y})\frac{\tilde{p}(\boldsymbol{\theta})p(\boldsymbol{y})}{p(\boldsymbol{\theta})\tilde{p}(\boldsymbol{y})} \tag{20}$$

The procedure for computing the reduced posterior is the following: first, we integrate over the parameters to obtain the evidence ratio of the two models:

$$\tilde{p}(\boldsymbol{y}) = p(\boldsymbol{y}) \int p(\boldsymbol{\theta}|\boldsymbol{y})\frac{\tilde{p}(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}d\boldsymbol{\theta} \tag{21}$$

Then, we define an approximate posterior $q(\boldsymbol{\theta})$ and we compute the reduced VFE in terms of the full model:

$$\mathcal{F}[\tilde{p}(\boldsymbol{\theta})] \approx \mathcal{F}[p(\boldsymbol{\theta})] + \ln \mathop{\mathbb{E}}_q \left[\frac{\tilde{p}(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}\right] \tag{22}$$

Correspondingly, the approximate posterior of the reduced model can be written in terms of the full model:

$$\ln \tilde{q}(\boldsymbol{\theta}) = \ln q(\boldsymbol{\theta}) + \ln \frac{\tilde{p}(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} - \ln \mathop{\mathbb{E}}_q \left[\frac{\tilde{p}(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}\right] \tag{23}$$

In order to apply this technique to the communication between discrete and continuous models, we first consider as parameters the hidden causes $\boldsymbol{v}$ and denote with $\boldsymbol{y}$ their continuous observations. We then assume that the full prior probability depends on a discrete outcome:

$$p(\boldsymbol{v}|\boldsymbol{o}) = \mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Pi}_v^{-1}) \tag{24}$$

In this way, the parameter $\boldsymbol{\eta}$ acts as a prior over the hidden causes. We also assume that the agent maintains M reduced models $\tilde{p}_m(\boldsymbol{v}, \boldsymbol{y})$, where each prior distribution is:

$$\tilde{p}_m(\boldsymbol{v}|\boldsymbol{o}) = \mathcal{N}(\boldsymbol{\eta}_m, \boldsymbol{\Pi}_m^{-1}) \tag{25}$$

These models correspond to specific hypotheses or assumptions that the agent believes to be possible causes of the sensorium. In other words, it is the reduced prior $\boldsymbol{\eta}_m$ that encodes a prior configuration for the continuous model, and each of them is associated with a particular discrete outcome $o_{\tau,m}$. The latter are computed by marginalizing $\boldsymbol{o}_{\pi,\tau}$ over all policies:

$$\boldsymbol{o}_\tau = \sum_\pi \pi_\pi \boldsymbol{o}_{\pi,\tau} \tag{26}$$

This means that the prior of the full model, which represents the actual agent's parameters of the current state of the world, is obtained by simply performing a Bayesian Model Average between every specified model:

$$\boldsymbol{\eta} = \sum_m o_{\tau,m} \boldsymbol{\eta}_m \tag{27}$$

thus transforming discrete probabilities into a continuous value. This quantity will then bias the belief over hidden causes $\boldsymbol{\nu}$ as in Equations 7-8. Similarly, having defined the full and reduced approximate posteriors:

$$\begin{aligned} q(\boldsymbol{v}) &= \mathcal{N}(\boldsymbol{\nu}, \boldsymbol{P}_v^{-1}) \\ \tilde{q}_m(\boldsymbol{v}) &= \mathcal{N}(\boldsymbol{\nu}_m, \boldsymbol{P}_m^{-1}) \end{aligned} \tag{28}$$

the ascending message that will act as a discrete observation for the higher-level model is computed through Bayesian Model Comparison, i.e., by comparing the probability that each discrete outcome model can be the true cause of the sensory input. In this case, the free energy computed in Equation 22 acts as a hint to how well a reduced model explains continuous observations, and it corresponds to the prior surprise plus the log-evidence for each outcome model accumulated over time:

$$\mathcal{F}_m = -\ln o_{\tau,m} - \int_0^T \mathcal{L}_m dt \tag{29}$$

The Laplace approximation [45] leads to a very simple form of the second RHS term:

$$\mathcal{L}_m = \frac{1}{2} \ln |\boldsymbol{\Pi}_m \boldsymbol{P}_v \boldsymbol{P}_m^{-1} \boldsymbol{\Sigma}_v| - \frac{1}{2}(\boldsymbol{\nu}^T \boldsymbol{P}_v \boldsymbol{\nu} - \boldsymbol{\nu}_m^T \boldsymbol{P}_m \boldsymbol{\nu}_m - \boldsymbol{\eta}^T \boldsymbol{\Pi}_v \boldsymbol{\eta} + \boldsymbol{\eta}_m^T \boldsymbol{\Pi}_m \boldsymbol{\eta}_m) \tag{30}$$

where:

$$\begin{aligned} \boldsymbol{\nu}_m &= \boldsymbol{P}_m^{-1}(\boldsymbol{P}_v \boldsymbol{\nu} - \boldsymbol{\Pi}_v \boldsymbol{\eta} + \boldsymbol{\Pi}_m \boldsymbol{\eta}_m) \\ \boldsymbol{P}_m &= \boldsymbol{P}_v - \boldsymbol{\Pi}_v + \boldsymbol{\Pi}_m \end{aligned} \tag{31}$$

After having computed the free energy associated with each model, they are normalized through a softmax function to get a proper probability, which is finally used to estimate the current discrete state:

$$\begin{aligned} \boldsymbol{r}_\tau &= \sigma(-\boldsymbol{\mathcal{F}}) \\ \boldsymbol{s}_{\pi,\tau} &= \sigma(\ln \boldsymbol{B}_{\pi,\tau-1} \boldsymbol{s}_{\pi,\tau-1} + \ln \boldsymbol{B}_{\pi,\tau}^T \boldsymbol{s}_{\pi,\tau+1} + \ln \boldsymbol{A}^T \boldsymbol{r}_\tau) \end{aligned} \tag{32}$$

Each value of $\boldsymbol{r}_\tau$ then indicates how probable is a discrete outcome based on the accumulated continuous evidence. Contrary to the discrete-only case, in a hybrid model the discrete actions are only implicitly used to compute the posterior over policies, and the link between the two frameworks is possible by associating each discrete outcome model $o_{\tau,m}$ to a particular continuous prior $\boldsymbol{\eta}_m$.

4.4 Dynamic Multi-step Tasks in Active Inference

To focus on the control aspects, we assume a simulated agent endowed with the following simplified sensory modalities: (i) a proprioceptive observation for the arm's joint angles (with dimension $n_p = 8$); (ii) a visual observation encoding the positions of hand, target, and goal (with dimension $n_v = 3$); (iii) the target's dimension; (iv) a tactile observation:

$$\boldsymbol{y} = \{\boldsymbol{y}_p, \boldsymbol{y}_v, y_d, y_t\} \tag{33}$$

Visual features may be inferred from a rich visual input [42], but for simplicity we directly provided the information needed in a high-level spatial domain, i.e., each component of $\boldsymbol{y}_v = [\boldsymbol{y}_{v,a}, \boldsymbol{y}_{v,t}, \boldsymbol{y}_{v,g}]$ is composed of two elements encoding the position of the hand, the target, and the goal in 2D Cartesian coordinates. Finally, tactile observations inform the agent through a Boolean function whether all the fingers – the last four joints – are touching an external object. Since it is here superfluous, we have only considered the Cartesian position of the hand, although a more realistic scenario should consist of information about all intermediate limb positions, which may be used to infer the correct posture from exteroceptive sensations only [36].

The agent also maintains latent states over all sensory modalities, so that the attractors only depend on what the agent believes is the current state of affairs of the world, and not on fixed states or sensory signals directly. In particular, since

the positions and dimensions of the objects are not static, a belief over bodily states is not sufficient, and we need to also encode information about both target and goal. Taking into account the nature of the task – which can be better tackled at the Cartesian level – we consider beliefs both in the intrinsic $\boldsymbol{\mu}_i$ and extrinsic $\boldsymbol{\mu}_e$ domains. As shown in the Supplementary Materials, it is convenient to express in both such domains also the possible objects that the agent may want to interact with. In this way, for each object the agent can automatically infer a possible joint configuration (i.e., affordance), which can be further restricted by some defined priors (e.g., depending on a particular grip).

Furthermore, the agent also maintains a belief over limb lengths $\boldsymbol{\mu}_l$ to compute the extrinsic coordinates of the hand and every object. This belief will not be updated by inference, although it is possible to learn it through a hierarchical model [36]. Finally, we also keep a belief over tactile observations μ_t. Summing up:

$$\boldsymbol{\mu} = \{\boldsymbol{\mu}_i, \boldsymbol{\mu}_e, \boldsymbol{\mu}_l, \mu_d, \mu_t\} \tag{34}$$

where the first two beliefs comprise components corresponding to the arm, target, and goal, reflecting the decomposition of the visual sensory signal:

$$\boldsymbol{\mu}_i = [\boldsymbol{\mu}_{i,a} \quad \boldsymbol{\mu}_{i,t} \quad \boldsymbol{\mu}_{i,g}] \qquad\qquad \boldsymbol{\mu}_e = [\boldsymbol{\mu}_{e,a} \quad \boldsymbol{\mu}_{e,t} \quad \boldsymbol{\mu}_{e,g}] \tag{35}$$

Since joint angles generate extrinsic coordinates with a one-to-one mapping, it is simpler to place the intrinsic state-space at the top of this hierarchy, and define a kinematic generative model $\boldsymbol{g}_e$ as:

$$\boldsymbol{p}_e = \boldsymbol{g}_e(\boldsymbol{\mu}_i) = [\boldsymbol{J}(\boldsymbol{\mu}_{i,a}, \boldsymbol{\mu}_l) \quad \boldsymbol{J}(\boldsymbol{\mu}_{i,t}, \boldsymbol{\mu}_l) \quad \boldsymbol{J}(\boldsymbol{\mu}_{i,g}, \boldsymbol{\mu}_l)] + \boldsymbol{w}_e \tag{36}$$

Here, $\boldsymbol{p}_e$ is the extrinsic prediction, $\boldsymbol{J}(\boldsymbol{\theta}, \boldsymbol{l})$ is the forward model:

$$\boldsymbol{J}(\boldsymbol{\theta}, \boldsymbol{l}) = \begin{bmatrix} l_1 c_1 + l_2 c_{12} + l_3 c_{123} + (l_4 + l^*) c_{1234} \\ l_1 s_1 + l_2 s_{12} + l_3 s_{123} + (l_4 + l^*) s_{1234} \end{bmatrix} \tag{37}$$

where l_n is the length of the n-th segment, and we used a compact notation to indicate the sine and cosine of the sum of angles. Since we are not interested in the fingers' positions during the reaching task, the kinematic generative model only computes the Cartesian position of the hand, which is found by extending the length of the last limb by a grasping distance l^*. Similarly, proprioceptive predictions are generated through a function $\boldsymbol{g}_p$, which is a simple mapping that extracts the arm's joint angles from the intrinsic belief:

$$\boldsymbol{p}_p = \boldsymbol{g}_p(\boldsymbol{\mu}_i) = \boldsymbol{\mu}_{i,a} + \boldsymbol{w}_p \tag{38}$$

while exteroceptive sensations – here, Cartesian position, dimension and touch – are generated from the following functions:

$$\boldsymbol{p}_v = \boldsymbol{g}_v(\boldsymbol{\mu}_e) = \boldsymbol{\mu}_e + \boldsymbol{w}_v \qquad p_d = g_d(\mu_d) = \mu_d + w_d \qquad p_t = g_t(\mu_t) = \mu_t + w_t \tag{39}$$

All the predictions are sampled from Gaussian probability distributions:

$$\begin{aligned} &\mathcal{N}(\boldsymbol{\mu}_e | \boldsymbol{g}_e(\boldsymbol{\mu}_i), \pi_e^{-1}) \\ &\mathcal{N}(\boldsymbol{y}_p | \boldsymbol{g}_p(\boldsymbol{\mu}_i), \pi_p^{-1}) \\ &\mathcal{N}(\boldsymbol{y}_v | \boldsymbol{g}_v(\boldsymbol{\mu}_e), \pi_v^{-1}) \end{aligned} \qquad\qquad \begin{aligned} &\mathcal{N}(y_d | g_d(\mu_d), \pi_d^{-1}) \\ &\mathcal{N}(y_t | g_t(\mu_t), \pi_t^{-1}) \end{aligned} \tag{40}$$

where π_e, π_p, π_v, π_d and π_t are their precisions. The prediction errors are finally computed by comparing predictions with the corresponding observations:

$$\begin{aligned} \boldsymbol{\varepsilon}_e &= \boldsymbol{\mu}_e - \boldsymbol{g}_e(\boldsymbol{\mu}_i) \\ \boldsymbol{\varepsilon}_p &= \boldsymbol{y}_p - \boldsymbol{g}_p(\boldsymbol{\mu}_i) \\ \boldsymbol{\varepsilon}_v &= \boldsymbol{y}_v - \boldsymbol{g}_v(\boldsymbol{\mu}_e) \end{aligned} \qquad\qquad \begin{aligned} \varepsilon_d &= y_d - g_d(\mu_d) \\ \varepsilon_t &= y_t - g_t(\mu_t) \end{aligned} \tag{41}$$

The multi-step task described can be solved in two different ways: by combining the continuous and discrete frameworks, or by using a continuous-only model. The dynamic nature of the task can be tackled through the definition of intentions [42], defined as functions that generate a possible future state from the current continuous belief. What differs between the two approaches is where such intentions are embedded and how they affect the system dynamics.

The belief dynamics can take into account many factors including friction, gravity, etc. – but in the following we assume that it depends on a combination of such intentions. For the task considered, it is useful to define an intention for each step: (i) reach the target; (ii) reach the goal position; (iii) close the fingers to grasp the target; (iv) open the fingers to release the target. The first two intentions can be interchangeably realized in an intrinsic reference frame that sets the arm's joint angles equal to the inferred target/goal configuration:

$$\begin{aligned} \boldsymbol{\mu}_{i,0}^* &= \boldsymbol{i}_{i,0}(\boldsymbol{\mu}_i) = [\boldsymbol{\mu}_{i,t} \quad \boldsymbol{\mu}_{i,t} \quad \boldsymbol{\mu}_{i,g}] \\ \boldsymbol{\mu}_{i,1}^* &= \boldsymbol{i}_{i,1}(\boldsymbol{\mu}_i) = [\boldsymbol{\mu}_{i,g} \quad \boldsymbol{\mu}_{i,t} \quad \boldsymbol{\mu}_{i,g}] \end{aligned} \tag{42}$$

or at the extrinsic level, by setting the wrist position to the desired value:

$$\begin{aligned}\boldsymbol{\mu}^*_{e,0} &= \boldsymbol{i}_{e,0}(\boldsymbol{\mu}_e) = [\boldsymbol{\mu}_{e,t} \quad \boldsymbol{\mu}_{e,t} \quad \boldsymbol{\mu}_{e,g}] \\ \boldsymbol{\mu}^*_{e,1} &= \boldsymbol{i}_{e,1}(\boldsymbol{\mu}_e) = [\boldsymbol{\mu}_{e,g} \quad \boldsymbol{\mu}_{e,t} \quad \boldsymbol{\mu}_{e,g}]\end{aligned} \tag{43}$$

The grasping/placing intentions are instead accomplished by generating a future intrinsic belief with the fingers in a closed/open configuration:

$$\begin{aligned}\boldsymbol{\mu}^*_{i,2} &= \boldsymbol{i}_{i,2}(\boldsymbol{\mu}_i) = \left[\boldsymbol{h}^*_{i,2,a} \quad \boldsymbol{\mu}_{i,t} \quad \boldsymbol{\mu}_{i,h}\right] \\ \boldsymbol{\mu}^*_{i,3} &= \boldsymbol{i}_{i,3}(\boldsymbol{\mu}_i) = \left[\boldsymbol{h}^*_{i,3,a} \quad \boldsymbol{\mu}_{i,t} \quad \boldsymbol{\mu}_{i,h}\right]\end{aligned} \tag{44}$$

where:

$$\begin{aligned}\boldsymbol{h}^*_{i,2,a} &= [\mu_{i,a,1} \quad \dots \quad \mu_{i,a,4} \quad 0 \quad 0 \quad -\theta_c \quad \theta_c] \\ \boldsymbol{h}^*_{i,3,a} &= [\mu_{i,a,1} \quad \dots \quad \mu_{i,a,4} \quad \theta_o \quad -\theta_o \quad 0 \quad 0]\end{aligned} \tag{45}$$

and θ_c and θ_o are the closed/open angles, which may further depend on the belief over target's dimension μ_d.

There are thus two sets of possible competing actions. Note also that if $\boldsymbol{i}_{i,0}$ is used for the whole trial so that the target is always linked to the wrist at the intrinsic level, the goal-reaching movement can be alternatively realized by only setting the target to the goal position through the following intention:

$$\boldsymbol{\mu}^*_{e,2} = \boldsymbol{i}_{e,2}(\boldsymbol{\mu}_e) = [\boldsymbol{\mu}_{e,a} \quad \boldsymbol{\mu}_{e,g} \quad \boldsymbol{\mu}_{e,g}] \tag{46}$$

4.4.1 Hybrid Models and Discrete Goals

Designing a discrete model involves the specification of the discrete states, a likelihood matrix, a transition matrix, a prior matrix, a preference matrix, and the actions. In the task considered, the hidden discrete states encode: (i) whether the hand and target are in the same position, at different positions, or both at the goal position; and (ii) whether the hand is open, closed or has grasped the target. In total, these factors combine in 9 possible process states. There are two likelihood matrices: (i) a matrix $\boldsymbol{A}$ that performs a simple identity mapping and will be used to obtain the discrete outcome for the continuous model; and (ii) a matrix $\boldsymbol{A}_t$ that returns a discrete tactile observation $\boldsymbol{o}_t$ – which is encoded by a Bernoulli distribution – signaling whether or not the target is grasped (hence, the continuous observation y_t is not used here). The transition matrix $\boldsymbol{B}$ is defined such that the target can be grasped only when both hand and target are in the same position. Note that every state has a certain probability – depending on the continuous sampling time – that the transition might fail. Finally, the discrete actions $\boldsymbol{u}$ correspond to the intentions defined above, with the addition of a null action needed for correctly computing the EFE at the goal position. The agent starts with the hand closed and not at the target's position. The trial is considered successful only if the agent has placed the target at the goal position and the hand is open.

Each discrete outcome has to be associated with a reduced continuous prior. The latter is usually considered static, but since we are dealing with a dynamic environment we let it depend on the intentions defined before. This has the advantage that if the reduced priors are generated at each discrete time step, a correct configuration over the hidden states can be imposed even if an object is moving – whose position will be dynamically inferred by the continuous model. However, note that a more realistic design would be to link the reduced priors to the hidden causes and let the latter encode the intentions.

Since the latter are defined in both intrinsic and extrinsic reference frames, a single discrete outcome $o_{\tau,m}$ generates two different sets of reduced priors, $\boldsymbol{\eta}_{i,m}$ and $\boldsymbol{\eta}_{e,m}$. For example, if the hand is open and at the target's position, the reduced priors are:

$$\boldsymbol{\eta}_{i,0} = \boldsymbol{i}_{i,2}(\boldsymbol{i}_{i,0}(\boldsymbol{\mu}_i)) \qquad\qquad \boldsymbol{\eta}_{e,0} = \boldsymbol{i}_{e,0}(\boldsymbol{\mu}_e) \tag{47}$$

Note that some outcomes generate the same reduced priors (e.g., the open hand condition does not affect the extrinsic prior), or do not impose any bias over the hidden states (e.g., if the hand and the target are at different positions, the mapping of the extrinsic level reduces to an identity and the corresponding reduced prior will be equal to the posterior). The full priors $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_e$ are then computed through a Bayesian Model Average:

$$\boldsymbol{\eta}_i = \sum_m o_{\tau,m} \cdot \boldsymbol{\eta}_{i,m} \qquad\qquad \boldsymbol{\eta}_e = \sum_m o_{\tau,m} \cdot \boldsymbol{\eta}_{e,m} \tag{48}$$

On the other side, ascending messages are computed by comparing the models of both intrinsic and extrinsic posteriors:

$$\tilde{\mathcal{F}}_m = -\ln o_{\tau,m} - \delta_i \sum_0^T \mathcal{L}_{m,i} dt - \delta_e \sum_0^T \mathcal{L}_{m,e} dt \tag{49}$$

where $\mathcal{L}_{m,i}$ and $\mathcal{L}_{m,e}$ are the log-evidences of the intrinsic and extrinsic beliefs respectively weighted by the gains δ_i and δ_e. The latter can be seen as quantities that modulate the magnitude of all the precisions involving a specific domain. Here, they were kept fixed throughout the tasks but they are in line with the hypothesis that higher values result in faster reaction times and evidence accumulation during perceptual decision-making [46].

The last thing to note is that the current state is inferred by taking into account the likelihoods of both the ascending message and the tactile observation – which is fundamental to infer whether the target has been correctly grasped:

$$\boldsymbol{s}_{\pi,\tau} = \sigma(\ln \boldsymbol{B}_{\pi,\tau-1}\boldsymbol{s}_{\tau-1} + \ln \boldsymbol{A}^T\boldsymbol{r}_\tau + \ln \boldsymbol{A}_t^T\boldsymbol{o}_t) \tag{50}$$

Having set up everything, all that remains is to compute the prior errors of the hidden states:

$$\boldsymbol{\varepsilon}_{\eta,i} = \boldsymbol{\mu}_i - \boldsymbol{\eta}_i \qquad\qquad \boldsymbol{\varepsilon}_{\eta,e} = \boldsymbol{\mu}_e - \boldsymbol{\eta}_e \tag{51}$$

The updates for the hidden states are then:

$$\begin{aligned}
\dot{\boldsymbol{\mu}}_i &= -\partial_{\boldsymbol{\mu}_i}\mathcal{F} = \partial\boldsymbol{g}_p^T\pi_p\boldsymbol{\varepsilon}_p + \partial\boldsymbol{g}_e^T\pi_e\boldsymbol{\varepsilon}_e - \pi_{\eta,i}\boldsymbol{\varepsilon}_{\eta,i} \\
\dot{\boldsymbol{\mu}}_e &= -\partial_{\boldsymbol{\mu}_e}\mathcal{F} = \partial\boldsymbol{g}_v^T\pi_v\boldsymbol{\varepsilon}_v - \pi_e\boldsymbol{\varepsilon}_e - \pi_{\eta,e}\boldsymbol{\varepsilon}_{\eta,e} \\
\dot{\mu}_d &= -\partial_{\mu_d}\mathcal{F} = \partial g_d^T\pi_d\varepsilon_d
\end{aligned} \tag{52}$$

where $\mathcal{F}$ is the VFE of the continuous model, while $\pi_{\eta,i}$ and $\pi_{\eta,e}$ are the precisions of the priors. Here, the gradient $\partial\boldsymbol{g}_e$ performs a kinematic inversion from intrinsic to extrinsic coordinates [36]:

$$\partial_{\boldsymbol{\mu}_i}\boldsymbol{g}_e = [\partial\boldsymbol{J}(\boldsymbol{\mu}_{i,a},\boldsymbol{\mu}_l) \quad \partial\boldsymbol{J}(\boldsymbol{\mu}_{i,t},\boldsymbol{\mu}_l) \quad \partial\boldsymbol{J}(\boldsymbol{\mu}_{i,g},\boldsymbol{\mu}_l)] \tag{53}$$

where:

$$\partial\boldsymbol{J}(\boldsymbol{\theta},\boldsymbol{l}) = \begin{bmatrix} -l_1s_1 - l_2s_{12} - l_3s_{123} - (l_4+l^*)s_{1234} \\ l_1c_1 + l_2c_{12} + l_3c_{123} + (l_4+l^*)c_{1234} \end{bmatrix} \tag{54}$$

In summary, the discrete model first computes policy probabilities depending on prior preferences, then averages them to obtain a discrete outcome. The latter is used to compute – through flexible intentions – the reduced priors, which are further weighted to get a final configuration that the continuous models have to realize. This configuration acts as a prior over the 0th order of the hidden states, and the belief update reduces to the computation of sensory gradients and prediction errors.

4.4.2 Continuous Model and Flexible Intentions

Whether or not a discrete model is necessary depends on the nature of the task, even when considering movements composed of different steps. In some cases, the agent knows that an a-priori defined sequence of actions (i.e., a habitual behavior) is all that is needed to solve a particular task, e.g., in rhythmic movements or, as in this case, a simple pick and place operation. Such scenarios may present little uncertainty regarding the order of elementary actions that does not necessarily involve repeated online decision-making. As we show below, these kinds of multi-step tasks can be easily tackled by a continuous-only model, with a few particular advantages with respect to the first approach.

From the point of view of the continuous model, the mapping from discrete outcomes to reduced priors is fixed, and what changes from one time step to another are the specific model probabilities that generate different full priors. Therefore, we can exploit the stability of the environment – composed of repeated sequences of the same steps executed in the same order – and define hidden causes $\boldsymbol{\nu}_i$ and $\boldsymbol{\nu}_e$ for each intention of both modalities. Next, we decompose the prior distribution over the generalized beliefs $\tilde{\boldsymbol{\mu}}_i$ and $\tilde{\boldsymbol{\mu}}_e$ (in this case up to the 1st order) into different functions for each hidden cause:

$$p(\tilde{\boldsymbol{\mu}}_i) = p(\boldsymbol{\mu}_i)\prod_j p(\boldsymbol{\mu}'_{i,j}|\boldsymbol{\mu}_i) \qquad\qquad p(\tilde{\boldsymbol{\mu}}_e) = p(\boldsymbol{\mu}_e)\prod_k p(\boldsymbol{\mu}'_{e,k}|\boldsymbol{\mu}_e) \tag{55}$$

where:

$$p(\boldsymbol{\mu}'_{i,j}|\boldsymbol{\mu}_i) = \mathcal{N}(\boldsymbol{\mu}'_{i,j}|\boldsymbol{f}_{i,j}(\boldsymbol{\mu}_i,\nu_{i,j}),\gamma_{i,j}^{-1}) \qquad\qquad p(\boldsymbol{\mu}'_{e,k}|\boldsymbol{\mu}_e) = \mathcal{N}(\boldsymbol{\mu}'_{e,k}|\boldsymbol{f}_{e,k}(\boldsymbol{\mu}_e,\nu_{e,j}),\gamma_{e,k}^{-1}) \tag{56}$$

and $\gamma_{i,j}$ and $\gamma_{e,k}$ are the corresponding precisions. Each function acts as an attractive force proportional to the error between the current and desired beliefs:

$$\begin{aligned}
\boldsymbol{f}_{i,j}(\boldsymbol{\mu}_i,v_{i,j}) &= v_{i,j}\boldsymbol{e}_{i,j} \qquad & \boldsymbol{e}_{i,j} &= \boldsymbol{\mu}^*_{i,j} - \boldsymbol{\mu}_i \\
\boldsymbol{f}_{e,k}(\boldsymbol{\mu}_e,v_{e,j}) &= v_{e,j}\boldsymbol{e}_{e,k} \qquad & \boldsymbol{e}_{e,k} &= \boldsymbol{\mu}^*_{e,k} - \boldsymbol{\mu}_e
\end{aligned} \tag{57}$$

where $\boldsymbol{\mu}^*_{i,j}$ and $\boldsymbol{\mu}^*_{e,k}$ are the future states computed through the intentions defined above. Note that the hidden causes can either act as attractor gain, or specify the relative strength of one state over the other [42, 47], which here means that they can be modulated to achieve a multi-step behavior. In fact, since the belief dynamics already stores and embeds every future goal, the belief will follow the contribution of all active states. In particular, we want the following behavior: (i) until the hand has not reached the target, execute only $\boldsymbol{i}_{e,0}$ (i.e., reach the target) and $\boldsymbol{i}_{i,3}$ (i.e., open the hand); (ii) as soon as the hand reaches the target, $\boldsymbol{i}_{i,2}$ should start (i.e., close the hand); (iii) when the agent has grasped the target, the target-reaching intention should be replaced by the $\boldsymbol{i}_{e,1}$ (i.e., move the hand toward the goal position along with the target); (iv) when the agent has reached the goal position, execute $\boldsymbol{i}_{i,3}$ again to release the target. Note that the target-reaching intention should be active until the agent has successfully grasped the target, since the latter may be moving and the grasping intention might fail. The desired behavior can be easily implemented by computing and combining Boolean functions (or sigmoid functions to have smooth transitions between states) of the beliefs. Having defined the dynamics prediction errors for each intention:

$$\boldsymbol{\varepsilon}_{\mu,i,j} = \boldsymbol{\mu}'_i - \boldsymbol{\mu}'_{i,j} \qquad\qquad \boldsymbol{\varepsilon}_{\mu,e,k} = \boldsymbol{\mu}'_e - \boldsymbol{\mu}'_{e,k} \tag{58}$$

the update for the intrinsic and extrinsic beliefs will be a combination of sensory prediction errors (Equation 41) and attractor dynamics:

$$\begin{aligned}
\dot{\tilde{\boldsymbol{\mu}}}_i &= \mathcal{D}\tilde{\boldsymbol{\mu}}_i - \partial_{\tilde{\boldsymbol{\mu}}_i}\mathcal{F} = \begin{bmatrix} \boldsymbol{\mu}'_i + \partial \boldsymbol{g}_p^T \pi_p \boldsymbol{\varepsilon}_p + \partial \boldsymbol{g}_e^T \pi_e \boldsymbol{\varepsilon}_e + \sum \partial \boldsymbol{f}_{i,j}^T \gamma_{i,j} \boldsymbol{\varepsilon}_{\mu,i,j} \\ - \sum \gamma_{i,j} \boldsymbol{\varepsilon}_{\mu,i,j} \end{bmatrix} \\
\dot{\tilde{\boldsymbol{\mu}}}_e &= \mathcal{D}\tilde{\boldsymbol{\mu}}_e - \partial_{\tilde{\boldsymbol{\mu}}_e}\mathcal{F} = \begin{bmatrix} \boldsymbol{\mu}'_e + \partial \boldsymbol{g}_v^T \pi_v \boldsymbol{\varepsilon}_v - \pi_e \boldsymbol{\varepsilon}_e + \sum \partial \boldsymbol{f}_{e,k}^T \gamma_{e,k} \boldsymbol{\varepsilon}_{\mu,e,k} \\ - \sum \gamma_{e,k} \boldsymbol{\varepsilon}_{\mu,e,k} \end{bmatrix}
\end{aligned} \tag{59}$$

Finally, we update the beliefs over target's dimension and tactile sensation using the corresponding prediction errors:

$$\dot{\mu}_d = \partial g_d^T \pi_d \varepsilon_d \qquad\qquad \dot{\boldsymbol{\mu}}_t = \partial \boldsymbol{g}_t^T \pi_t \boldsymbol{\varepsilon}_t \tag{60}$$

4.4.3 An in-depth comparison

Comparing the continuous belief update with the discrete state inference, we can note a relationship between the policy-dependent $\boldsymbol{s}_{\pi,\tau}$ states computed through the transition matrices, and the trajectories $\boldsymbol{\mu}'_{i,j}$ and $\boldsymbol{\mu}'_{e,k}$ computed through the continuous intentions (see Figure 3b). While a policy-independent discrete state $\boldsymbol{s}_\tau$ is found by calculating a Bayesian Model Average with each policy probability π_π, in the continuous case the hidden states $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_e$ are computed by averaging each trajectory with the corresponding hidden causes $\nu_{i,j}$ and $\nu_{e,k}$ at each step of the task.

If we consider the hybrid model, we can also note a similarity between the probability distributions of the reduced priors of Equation 25 and the probabilities of Equation 56 that compose the dynamics function. In the former case, each reduced prior is averaged through each outcome model probability $o_{\tau,m}$. In this case, however, the full prior biases the hidden states through the overall prediction error $\varepsilon_{\eta,i}$, which already contains the final configuration that the continuous model must realize until the next discrete step. In the continuous-only model, the intentions are instead used to compute different directions of update that independently pull the belief over the hidden states. Also, the prior precision $\pi_{\eta,i}$ of Equation 52 and the dynamics precisions $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_e$ of Equation 59 play the same role of encoding the strength of the intention dynamics – that steers the belief toward a desired state – with respect to the sensory likelihood – that keeps the agent close to what it is currently perceiving. Crucially, while the weighted prediction error of the hybrid model is:

$$\pi_{\eta,i}(\boldsymbol{\mu}_i - \sum o_{\tau,m} \boldsymbol{\eta}_{i,m}) \tag{61}$$

if the precisions have the same value, the combination of the weighted dynamics prediction error of the continuous-only model takes the form:

$$\boldsymbol{\mu}'_i + \gamma_i(\boldsymbol{\mu}_i - \sum \nu_{i,j} \boldsymbol{h}_{i,j}) \tag{62}$$

However, note that encoding different precisions for each intention permits an additional modulation that can be seen as the level of agent's confidence for a particular task.

Finally, note that in the hybrid model the extrinsic belief $\boldsymbol{\mu}_e$ can be biased through two different pathways: directly from the reduced prior of the discrete model, or indirectly through forward kinematics of the intrinsic belief $\boldsymbol{\mu}_i$. Correspondingly, the latter can be changed either by imposing a particular configuration in the intrinsic domain through the discrete model, or by the backward pathway of the extrinsic belief, through computation of the gradient $\partial \boldsymbol{g}_e$ that performs a kinematic inversion [36]. The same holds for the continuous-only model, where the hidden states are biased by their own dynamics instead of the full priors.

Acknowledgments

This research received funding from the European Union's Horizon H2020-EIC-FETPROACT-2019 Programme for Research and Innovation under Grant Agreement 951910 to I.S., the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 to GP, the European Research Council under the Grant Agreement No. 820213 to GP, and the Italian Ministry for Research MIUR under Grant Agreement PRIN 2017KZNZLN to I.S. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Thomas Parr, Rajeev Vijay Rikhye, Michael M. Halassa, and Karl J. Friston. Prefrontal Computation as Active Inference. *Cerebral Cortex*, 30(2):682–695, 2020.

[2] Karl J. Friston, Richard Rosch, Thomas Parr, Cathy Price, and Howard Bowman. Deep temporal models and active inference. *Neuroscience and Biobehavioral Reviews*, 77(November 2016):388–402, 2017.

[3] Daniel E Callan and Eiichi Naito. from Expert and Novice Athletes. 27(4):183–188, 2014.

[4] Francesco Di Russo, Sabrina Pitzalis, Teresa Aprile, and Donatella Spinelli. Effect of practice on brain activity: An investigation in top-level rifle shooters. *Medicine and Science in Sports and Exercise*, 37(9):1586–1593, 2005.

[5] F. Fattapposta, G. Amabile, M. V. Cordischi, D. Di Venanzio, A. Foti, F. Pierelli, C. D'Alessio, F. Pigozzi, A. Parisi, and C. Morrocutti. Long-term practice effects on a new skilled motor learning: An electrophysiological study. *Electroencephalography and Clinical Neurophysiology*, 99(6):495–507, 1996.

[6] Lucio Marinelli, Angelo Quartarone, Mark Hallett, Giuseppe Frazzitta, and Maria Felice Ghilardi. The many facets of motor learning and their relevance for Parkinson's disease. *Clinical Neurophysiology*, 128(7):1127–1141, 2017.

[7] Elisenda Bueichekú, Anna Miró-Padilla, María Ángeles Palomar-García, Noelia Ventura-Campos, María Antonia Parcet, Alfonso Barrós-Loscertales, and César Ávila. Reduced posterior parietal cortex activation after training on a visual search task. *NeuroImage*, 135:204–213, 2016.

[8] Alice Nieuwboer, Lynn Rochester, Liesbeth Müncks, and Stephan P. Swinnen. Motor learning in Parkinson's disease: limitations and potential for rehabilitation. *Parkinsonism and Related Disorders*, 15(SUPPL. 3):53–58, 2009.

[9] Giovanni Pezzulo, Francesco Rigoli, and Karl J. Friston. Hierarchical Active Inference: A Theory of Motivated Control. *Trends in Cognitive Sciences*, 22(4):294–306, 2018.

[10] Ann M. Graybiel. Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, 31:359–387, 2008.

[11] Reza Shadmehr and Henry H. Holcomb. Neural correlates of motor memory consolidation. *Science*, 277(5327):821–825, 1997.

[12] D M Wolpert, Z Ghahramani, and M Jordan. An internal model for sensorimotor integration - Wolpert et al. (1995).pdf. *Science*, 269(5):1880–1882, 1995.

[13] Karl Friston and Stefan Kiebel. Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521):1211–1221, 2009.

[14] Karl J. Friston, Jean Daunizeau, James Kilner, and Stefan J. Kiebel. Action and behavior: A free-energy formulation. *Biological Cybernetics*, 102(3):227–260, 2010.

[15] Thomas Parr, Giovanni Pezzulo, and Karl J Friston. *Active inference: the free energy principle in mind, brain, and behavior*. Cambridge, MA: MIT Press, 2021.

[16] Rajesh P.N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

[17] Karl Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010.

[18] Rick A. Adams, Stewart Shipp, and Karl J. Friston. Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, 218(3):611–643, 2013.

[19] Lancelot Da Costa, Thomas Parr, Noor Sajid, Sebastijan Veselic, Victorita Neacsu, and Karl Friston. Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, 99, 2020.

[20] Ryan Smith, Karl J. Friston, and Christopher J. Whyte. A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, 107:102632, 2022.

[21] Raphael Kaplan and Karl J. Friston. Planning and navigation as active inference. *Biological Cybernetics*, 112(4):323–343, 2018.

[22] Planning as inference. *Trends in Cognitive Sciences*, 16(10):485–488, 2012.

[23] Marc Toussaint. Probabilistic inference as a model of planned behavior. *Künstliche Intelligenz*, 3/09:23–29, 2009.

[24] Sherin Grimbergen, S S Grimbergen, · C Van Hoof, · P Mohajerin Esfahani, and · M Wisse. Active Inference for State Space Models: A Tutorial. 2019.

[25] Noor Sajid, Philip J. Ball, Thomas Parr, and Karl J. Friston. Active inference: demystified and compared. *Neural Computation*, 33(3):674–712, 2021.

[26] Philipp Schwartenbeck, Thomas H.B. Fitzgerald, Christoph Mathys, Ray Dolan, Martin Kronbichler, and Karl Friston. Evidence for surprise minimization over value maximization in choice behavior. *Scientific Reports*, 5:1–14, 2015.

[27] Philipp Schwartenbeck, Johannes Passecker, Tobias U. Hauser, Thomas H.B. Fitzgerald, Martin Kronbichler, and Karl J. Friston. Computational mechanisms of curiosity and goal-directed exploration. *eLife*, 8:1–45, 2019.

[28] Karl J. Friston, Jean Daunizeau, and Stefan J. Kiebel. Reinforcement learning or active inference? *PLoS ONE*, 4(7), 2009.

[29] Karl J Friston, S Samothrakis, and Read Montague. Active inference and agency: optimal control without cost functions. *Biological cybernetics*, (106):523–541, 2012.

[30] Karl J. Friston, Thomas Parr, and Bert de Vries. The graphical brain: Belief propagation and active inference. 1(4):381–414, 2017.

[31] Thomas Parr and Karl J. Friston. The Discrete and Continuous Brain: From Decisions to Movement—And Back Again. (September):2319–2347, 2018.

[32] Karl Friston, Thomas Parr, and Peter Zeidman. Bayesian model reduction. pages 1–32, 2018.

[33] Karl Friston and Will Penny. Post hoc Bayesian model selection. *NeuroImage*, 56(4):2089–2099, 2011.

[34] Alexander Tschantz, Laura Barca, Domenico Maisto, Christopher L. Buckley, Anil K. Seth, and Giovanni Pezzulo. Simulating homeostatic, allostatic and goal-directed forms of interoceptive control using active inference. *Biological Psychology*, 169, 2022.

[35] Janelle Weaver. Motor Learning Unfolds over Different Timescales in Distinct Neural Systems. *PLoS Biology*, 13(12):1–2, 2015.

[36] Matteo Priorelli, Giovanni Pezzulo, and Ivilin Peev Stoianov. Deep kinematic inference affords efficient and scalable control of bodily movements. *bioRxiv*, 2023.

[37] Emanuel Todorov. Optimality principles in sensorimotor control. *Nature Neuroscience*, 7:907–915, 2004.

[38] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[39] Jakob Hohwy. *The Predictive Mind*. Oxford University Press UK, 2013.

[40] Daniel Baldauf, He Cui, and Richard A. Andersen. The posterior parietal cortex encodes in parallel both goals for double-reach sequences. *Journal of Neuroscience*, 28(40):10081–10089, 2008.

[41] Stefan Schaal, Dagmar Sternad, Rieko Osu, and Mitsuo Kawato. Rhythmic arm movement is not discrete. *Nature Neuroscience*, 7(10):1137–1144, 2004.

[42] Matteo Priorelli and Ivilin Peev Stoianov. Flexible intentions: An active inference theory. *Front. Comput. Neurosci.*, Mar. 2023.

[43] Francesco Mannella, Federico Maggiore, Manuel Baltieri, and Giovanni Pezzulo. Active inference through whiskers. *Neural Networks*, 144:428–437, 2021.

[44] Marshall G. Shuler and Mark F. Bear. Reward Timing in the Primary Visual Cortex. *Science*, (March):1606–1610, 2015.

[45] Karl Friston, Jérémie Mattout, Nelson Trujillo-Barreto, John Ashburner, and Will Penny. Variational free energy and the Laplace approximation. *NeuroImage*, 34(1):220–234, 2007.

[46] Thomas H.B. FitzGerald, Rosalyn J. Moran, Karl J. Friston, and Raymond J. Dolan. Precision and neuronal dynamics in the human posterior parietal cortex during evidence accumulation. *NeuroImage*, 107:219–228, 2015.

[47] Matteo Priorelli and Ivilin Peev Stoianov. Intention Modulation for Multi-Step Tasks in Continuous Time Active Inference. In *Active Inference, Third International Workshop, IWAI 2022, Grenoble, France, Sept 19, 2022*, 2022.

5 Supplementary Materials

5.1 Simulating Affordances

Defining continuous intentions also in intrinsic coordinates has the advantage that high hierarchical levels can impose priors not only on the hand positions, but also on specific configurations of the arm and the fingers. This is useful when an agent wants to grasp an object in a particular manner depending on the task considered (e.g., power grip vs precision grip). Figure 7 shows a simple demonstration of this behavior: here, the agent is required to grasp the object with a counterclockwise wrist rotation, and then place it at the goal position with a clockwise rotation. The arm starts with the fingers closed and the target belief is initialized at the hand position. Then it starts reaching the target while slowly opening the hand and preparing the correct wrist rotation. Since it is subject to multi-domain constraints, this configuration is maintained until the moving target is grasped. Note that there is tradeoff between reaching the target and maintaining the imposed constraint on the wrist, as can be seen in the middle of the movement where the agent waits until the target can be grasped with the correct rotation even if the latter is already in its peripersonal space. When the target is grasped, it rotates the wrist again by 180 degrees clockwise and correctly places the target in the goal position.

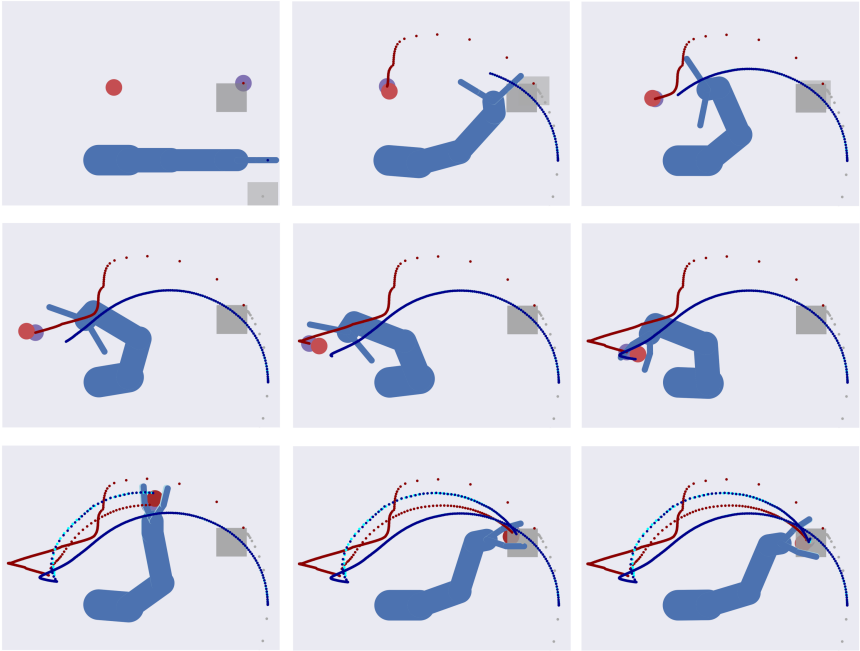


Figure 7: Sequence of time frames of the grasping task with affordances. Real and estimated target positions are represented by red and purple circles, real and estimated goal positions by grey and dark grey squares, and real and estimated limb positions in blue and cyan. Hand, target, and goal belief trajectories are represented respectively with a blue, red, and grey dotted line.

5.2 Supplementary Movie 1

Simulation of the dynamic pick-and-place task with a hybrid model.

5.3 Supplementary Movie 2

Simulation of the dynamic pick-and-place task with a continuous-only model. Also see Figure 4.

5.4 Supplementary Movie 3

Simulation of the dynamic pick-and-place task with affordances. Also see Figure 7.