Embodied decisions as active inference

Matteo Priorelli, Ivilin Peev Stoianov, Giovanni Pezzulo*

Abstract

Decision-making is often conceptualized as a serial process, during which sensory evidence is accumulated for the choice alternatives until a certain threshold is reached, at which point a decision is made and an action is executed. This *decide-then-act* perspective has successfully explained various facets of perceptual and economic decisions in the laboratory, in which action dynamics are usually irrelevant to the choice. However, living organisms often face another class of decisions – called *embodied decisions* – that require selecting between potential courses of actions to be executed timely in a dynamic environment, e.g., for a lion, deciding which gazelle to chase and how fast to do so. Studies of embodied decisions reveal two aspects of goal-directed behavior in stark contrast to the serial view. First, that decision and action processes can unfold in parallel; second, that action-related components, such as the motor costs associated with the choice alternatives and required to "change mind" between them, exert a feedback effect on the decision taken. Here, we show that these signatures of embodied decisions emerge naturally in active inference – a framework that simultaneously optimizes perception and action, according to the same (free energy minimization) imperative. We show that optimizing embodied choices requires a continuous feedback loop between motor planning (where beliefs about choice alternatives guide action dynamics) and motor inference (where action dynamics finesse beliefs about choice alternatives). Furthermore, our active inference simulations reveal the normative character of embodied decisions in ecological settings – namely, achieving an effective balance between a high accuracy and a low risk of losing valid opportunities.

1 Introduction

Decision-making is traditionally conceptualized as a serial, *decision-then-act* process, in which sensory evidence is accumulated until a certain threshold, at which point a decision is made and an action is executed. This approach, as formalized by *drift-diffusion* and related models, has been very useful in analyzing behavioral and neural data from laboratory studies of perceptual and economic decisions [1, 2, 3]. In these studies, participants select between fixed choice alternatives (usually two) reflecting perceptual judgments (e.g., motion discrimination) or economic offers (e.g., lotteries).

However, animals often face a different class of *embodied decisions*, which imply the choice between courses of actions to be immediately executed in dynamic environments; for example, for a lion, the choice of which gazelle to chase or for a soccer player, the choice of which teammate to which passing the ball [4, 5, 6, 7, 8, 9, 10]. To face with the demands of these embodied decisions, animals often need to specify, prepare and sometimes execute actions in parallel to the decision process – as captured by the notion of *affordance competition* [11, 5].

These considerations motivate a series of experiments that use continuous measures of performance during perceptual and economic choices; for example, track mouse of hand kinematics while participants moved from the start to the response buttons [12, 13]. Despite their simplicity, these experiments permit analyzing the dynamic processes leading to a decision and the reciprocal influences between

*Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

Preprint. Under review.

the ongoing deliberation and movement (i.e., *decide-while-acting* or *continuous decisions* [14]). They reveal that participants move and deliberate simultaneously; they generally start moving very early on, either toward a specific target or in the middle, if they are more uncertain; they often revisit their decisions in the middle of the trial, as apparent by the curvature of their movements; and they sometimes change mind between the targets, as evident by drastic changes of trajectory [15, 16, 17, 18]. These findings eschew serial models and are better explained by parallel [19] or continuous flow models [20] in which unfolding perceptual and decision processes concurrently drive the preparation and possibly the overt execution of one or more responses in parallel – meaning that movements during the task provide a continuous readout of the ongoing deliberation. From a normative perspective, parallel models provide a way to realize decisions faster, which is crucial to survival as it avoids the risk of losing valued opportunity – sometimes at the cost of reduced accuracy [21].

Crucially, some studies of embodied decisions reveal feedback effects of action dynamics to decision processes that were previously ignored in serial and even parallel decision-making models. For example, one recurring finding is that the motor costs associated with different choice alternatives influence perceptual and economic decisions. During ambiguous perceptual decisions [22, 23] and value-based decisions [24], participants show a bias to select the response choice associated with the less costly movement. Changes of mind during a perceptual task are less frequent if the costs associated with changing movement direction are greater, such as when response buttons are farther apart [25]. Similarly, during an economic task, changes of mind following a perturbation of the movement trajectory are sensitive to the current state (position and velocity) of the motor system and less frequent when counteracting the perturbation would be more costly [26].

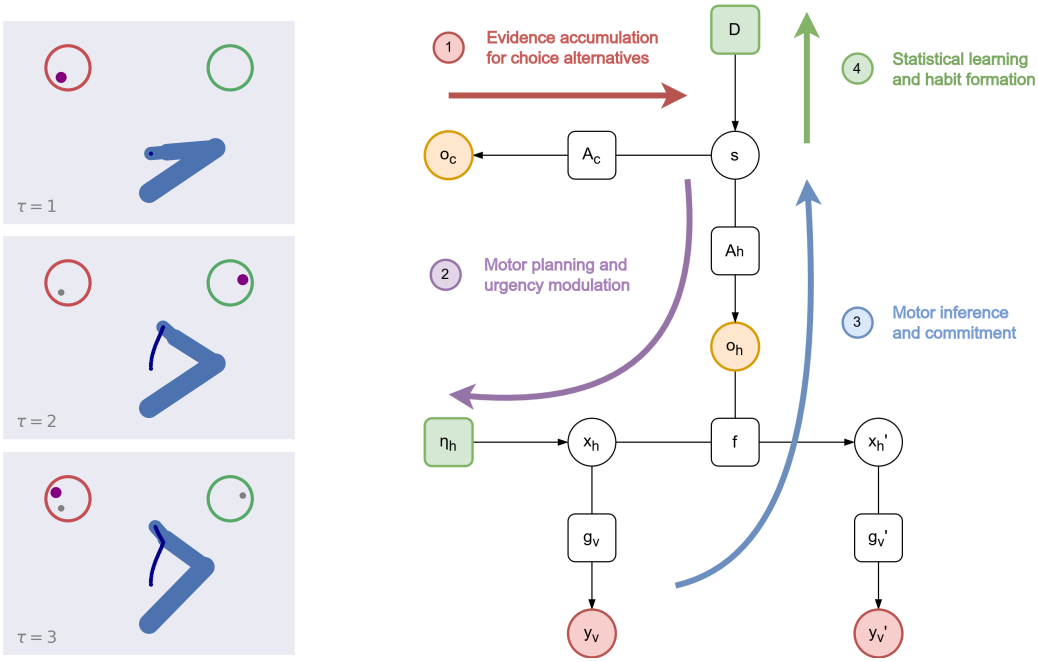
These and other studies suggest not only that deliberation continues after movement onset (in accordance with parallel models) but also that is affected by feedback from action dynamics (e.g., motor costs). This motivates a novel class of *embodied decision* models in which action is not the inert outcome of a decision process but influences it, forming a closed loop [27, 21]. These models are motivated by the fact that from an embodied perspective, the goal of the agent is not just selecting between choice alternatives (as in classical setups) but also simultaneously selecting between potential courses of action to reach the targets (often, within a deadline) and tracking the action itself – which means that both decision and action processes need to be jointly and continuously optimized. In turn, at the neural level, embodied decisions might require a *distributed consensus* across various brain networks that process outcome values and motor plans, rather than a centralized process as traditionally assumed [28].

Here, we show that the key signatures of embodied decisions emerge naturally in *active inference*, a framework that models perception and action selection as two aspects of the same objective of free energy minimization [29, 30, 31, 32]. By simulating embodied decisions as an active inference process, we are able to reproduce various empirical findings about the parallel unfolding of actions and decisions in time, as well as feedback effects of movement dynamics in perception. Furthermore, we illustrate the normative advantages of embodied choices over serial choices under time pressure.

2 Results

We illustrate the functioning of an active inference model of embodied decisions by simulating a two-alternative forced choice (2AFC) decision task with time-varying information, i.e., in which evidence for one choice or the other, expressed in terms of sequentially provided cues, changes throughout each trial, as in [33, 34] (Figure 1a). The agent has to move the 3-DoF arm from a start position (small blue dot at the center) to reach a left (red circle) or right (green circle) target button, to report which target has or will have more cues in it. During the task, 15 cues appear, one after the other, either in the left or the right circle, and then disappear leaving only the last one visible. The agent can start moving at any moment and the cues continue to appear normally during movement. The trial ends when the agent reaches one of the two buttons (or within a deadline).

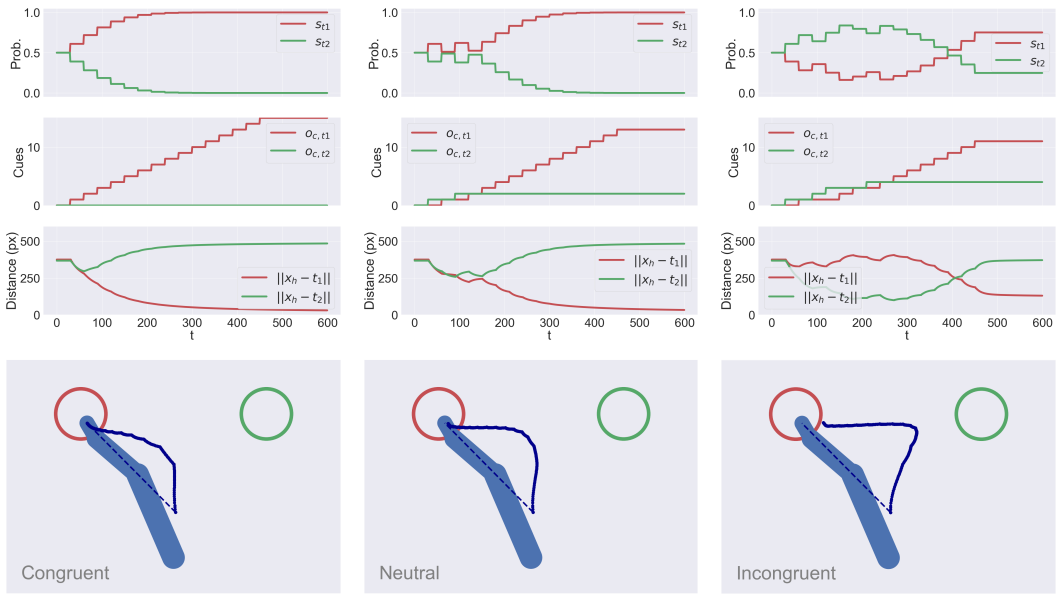
Crucially, by manipulating the sequence of cues, we compare the agent's decision dynamics in three conditions (or trial types): *congruent*, in which a greater proportion of cues initially appears in the correct target; *incongruent*, in which a greater proportion of cues initially appears in the incorrect target; and *neutral*, in which the proportion of cues initially appearing in both targets is balanced.



(a) (b)

Figure 1: Embodied decision setup and active inference model. (a) Experimental setup, during three consecutive discrete time steps τ. The agent controls a 3-DoF arm, which starts at a home position (blue dot) at an equal distance from the two targets (red and green circles). The current cue is represented with a big purple dot, while the old cues are represented with smaller grey dots. For each trial, the agent has to reach with the hand the target it believes will contain more circles. (b) Hybrid active inference model for embodied decisions. The model comprises four pathways, numbered from 1 to 4. In the first pathway, discrete hidden states $\boldsymbol{s}$, encoding the probability that each target is the correct choice for the current trial, are iteratively inferred by discrete cues $\boldsymbol{o}_c$ by inverting the cue likelihood matrix $\boldsymbol{A}_c$. In the second pathway, the hidden states $\boldsymbol{s}$ generate a particular combination of discrete hand dynamics $\boldsymbol{o}_h$ through the extrinsic likelihood matrix $\boldsymbol{A}_h$. Each hand dynamics $o_{h,m}$ is related to a continuous dynamics function $\boldsymbol{f}_m$, where the target positions are defined (see [35, 36] for more details). A forward message imposes a prior over the hand velocity (in Cartesian coordinates) $\boldsymbol{x}'_h$, while a backward message infers the related Cartesian position of the hand $\boldsymbol{x}_h$, ready for kinematic and dynamic inversions. In the third pathway, for each continuous time step t, the current position and velocity of the hand (in a visual domain) are inferred by continuous observations $\boldsymbol{y}_v$ and $\boldsymbol{y}'_v$, via the corresponding likelihood functions. For each discrete time step τ, the target probabilities $\boldsymbol{s}$ are also inferred by the current motor trajectory. Finally, in the fourth pathway, the prior $\boldsymbol{D}$ over the correct target is updated across trials, implementing habitual learning.

Empirical evidence indicates that these trial types significantly influence choice and movement dynamics. In the Tokens task [33], which is conceptually similar to this task (except for the fact that the old cues disappear), trial incongruency determines a greater number of errors and longer response times [33] and impacts the lateral hand position prior to a decision, when movement kinematics were recorded [37]. Similar results are observed in the Eriksen flanker task [34], which shares conceptual similarities with our task, because a correct visual target is presented along with either congruent or incongruent cues. The flanker task can be modeled as a progressive shift of attention toward the correct target over time, equivalent to sampling it from increasingly precise distributions [16]. This process explains the significant decrease of performance during incongruent conditions (*flanker effect*), because the initial sampling process is biased toward the incorrect response. The decrease in performance manifests as increased number of errors, reaction time, and curvature of the movement trajectory in tasks that tracked movement kinematics [38, 16]. Furthermore, measures of the motor potentials (ERPs: event-related brain potentials and EMG: electromyographic activity) after target presentation suggested that competing responses were activated simultaneously [39], in keeping with analogous findings through single-cell recordings of monkey dorsal premotor cortex [40].



(a) (b) (c)

Figure 2: Evidence accumulation during (a) congruent trials, (b) neutral trials, and (c) incongruent trials. The first row shows the dynamics of a sample trial for each condition: specifically, the first plot shows the discrete hidden states $\boldsymbol{s}$ encoding the two target probabilities over continuous time; the second plot shows the cumulative sum of the cue observations $\boldsymbol{o}_c$; the third plot shows the distances between the hand and the two targets. Note that the discrete signals are maintained for a whole discrete period τ, generating a stepped behavior. The second row shows the agent's average trajectory (in dark blue) across 100 trials for each condition. A dotted line of minimum distance between the initial hand position and the left target is also displayed.

Below, we show that a hybrid active inference model (i.e. a model composed of both discrete and continuous variables) that jointly optimizes decisions and actions reproduces these signatures of embodied choices. The model can be decomposed into four interacting pathways – evidence accumulation, motor planning, motor inference, and statistical learning across trials (along with habit formation) – see Figure 1b for a schematic illustration and Appendix B for technical details. Below we discuss these pathways and present simulations showing how they affect the agent's decision and action processes.

2.1 The first pathway: evidence accumulation for the choice alternatives

The first pathway is responsible for sequential evidence accumulation for the choice alternatives. It includes discrete hidden states $\boldsymbol{s}$, encoding the probability that each target is the correct choice for the current trial (i.e., the one that will contain the most cues). They are sampled from a categorical (here, binomial) distribution, i.e., $\boldsymbol{s} = Cat(\boldsymbol{D}) = [s_{t1} \quad s_{t2}]$, where $\boldsymbol{D}$ are the parameters of a Dirichlet distribution and define the agent's prior beliefs. In the following simulations, we initialize them with a uniform distribution for each trial. The discrete hidden states $\boldsymbol{s}$ generate two discrete predictions in parallel. The former – computed through the likelihood matrix $\boldsymbol{A}_c$ – is a prediction of the cue that the agent will observe next, with α_c playing the role of a weight factor, similar to the *drift rate* in drift-diffusion models. The latter – computed through the likelihood matrix $\boldsymbol{A}_h$ – corresponds to the hand dynamics: it predicts whether the hand will move toward the left target, toward the right target, or not move at all (with probability α_h). In short, the uncertainty of the two mappings affects how fast the evidence accumulation unfolds, and which movement strategy the agent adopts:

$$\boldsymbol{A}_c = \begin{bmatrix} 1-\alpha_c & \alpha_c \\ \alpha_c & 1-\alpha_c \end{bmatrix} \qquad \boldsymbol{A}_h = \begin{bmatrix} 1-\alpha_h & 0 & \alpha_h \\ 0 & 1-\alpha_h & \alpha_h \end{bmatrix} \tag{1}$$

Since the agent has to make a simple decision without planning in the future, we did not model discrete dynamics with transition matrices $\boldsymbol{B}$ (as defined in Appendix B). At each discrete step τ, a particular cue $\boldsymbol{o}_c$ and a particular hand dynamics $\boldsymbol{o}_h$ are observed and compared with the corresponding predictions. Hence, the inference of the discrete hidden states $\boldsymbol{s}$ follows the equation:

$$\boldsymbol{s}_\tau = \sigma_w(k_d \ln \boldsymbol{s}_{\tau-1} + \ln \boldsymbol{A}_c^T \boldsymbol{o}_{c,\tau} + k_h \ln \boldsymbol{A}_h^T \boldsymbol{o}_{h,\tau}) \tag{2}$$

where σ is a weighted softmax function:

$$\sigma_w(\boldsymbol{x}) = \frac{e^{wx_i}}{\sum_k h^{wx_k}} \tag{3}$$

with slope (or precision) w. A high value of w ensures fast transitions between discrete states, hence avoiding positions within the two targets. In short, the discrete update is a combination of a prior coming from the previous step (which is equal to $\boldsymbol{D}$ at the beginning of the trial) and two likelihoods. The first one contributes to sensory evidence accumulation and iteratively refines the choice, based on the sensory cues. The second likelihood links target estimation and hand dynamics; in this way, the variable $\boldsymbol{o}_h$ behaves as a sensory signal for the discrete model (similar to $\boldsymbol{o}_c$) and permits accumulating evidence from the agent's movements. We will unpack the role of the second likelihood in the next sections; here, to simulate the standard evidence accumulation, we let the second term depend on a parameter k_h, which we here set to 0 so that the inference of the correct choice only relies on sensory cues $\boldsymbol{o}_c$. Finally, we include a parameter k_d acting as a forgetting factor (which might be useful to deal with non-stationary tasks), which we keep fixed at 1 in our simulations. Also, we set $\alpha_h = 0$ (so that the agent starts moving at the beginning of the trial) and $\alpha_c = 0.4$.

We test the active inference model with the three conditions explained above. In *congruent* trials, cues move toward the correct target, with an initial probability of 80%, which then gradually increases and reaches 100% after 8 cues. In *neutral* and *incongruent* trials, the probabilities of the correct target are respectively initialized to 50% and 20% and then increase as in congruent trials. Each trial comprises 21 discrete time steps τ, each in turn comprising 30 continuous steps t. At $\tau = 0$, no cue is presented, but the agent can move. For the next 15 time steps, a cue per time step is presented. Finally, in the last 4 time steps, no cue is presented, but the agent can still move and reach the target.

Figure 2 shows that in all conditions, when movement onset is immediate the agent moves in between the two targets. In the *incongruent* condition, the agent first moves toward the wrong target and eventually changes mind. These results match qualitatively key empirical findings discussed in the introduction [15, 16, 17, 18]. To assess whether the model generates statistically different trajectories under the three conditions, we simulated 100 trials per condition and considered a widely used index of choice uncertainty: the *maximum deviation* of the trajectories from an ideal, straight line between the start and the correct target [12, 10]. We found significantly larger maximum deviation in *incongruent* (M = 125.97, SD = 25.0) compared to *neutral* (M = 88.19, SD = 30.73) trials and *neutral* compared to *congruent* (M = 56.64, SD = 18.56) trials (for all tests, $p < .001$).

2.2 The second pathway: motor planning and urgency

When facing the same task, different participant groups might show different strategies; for example, a conservative strategy to postpone movement until they feel sufficiently confident, or a risky strategy to guess the correct choice and start moving immediately [41].

In our model, the selection of a conservative versus risky strategy depends on two things. First, the weight of evidence accumulation, via the uncertainty parameter α_c: the lower the uncertainty, the faster the agent will form a strong belief about the correct target. Second, the movement *urgency*, via the uncertainty α_h over the likelihood matrix related to the hand dynamics: the lower α_h, the less certain the agent has to be about the correct target to start moving. Recall that $\boldsymbol{o}_h$ (the *discrete set of hand dynamics*) encodes the probability that the two targets generate a movement toward the left, toward the right, or no movement (i.e., $\boldsymbol{o}_h = [o_{h,t1}, o_{h,t2}, o_{h,s}]$).

Importantly, $\boldsymbol{o}_h$ specifies potential action plans, as it communicates with continuous hidden states specifying the instantaneous trajectory (position $\boldsymbol{x}_h$ and velocity $\boldsymbol{x}'_h$) of the agent's hand in extrinsic (e.g., Cartesian) coordinates. Specifically, each element of $\boldsymbol{o}_h$ is linked to a dynamics function:

$$\boldsymbol{f}(\boldsymbol{\mu}_h) = \begin{bmatrix} \boldsymbol{f}_{t1} \\ \boldsymbol{f}_{t1} \\ \boldsymbol{f}_s \end{bmatrix} = \begin{bmatrix} \lambda(\boldsymbol{p}_{t1} - \boldsymbol{\mu}_h) \\ \lambda(\boldsymbol{p}_{t2} - \boldsymbol{\mu}_h) \\ \boldsymbol{0} \end{bmatrix} \tag{4}$$

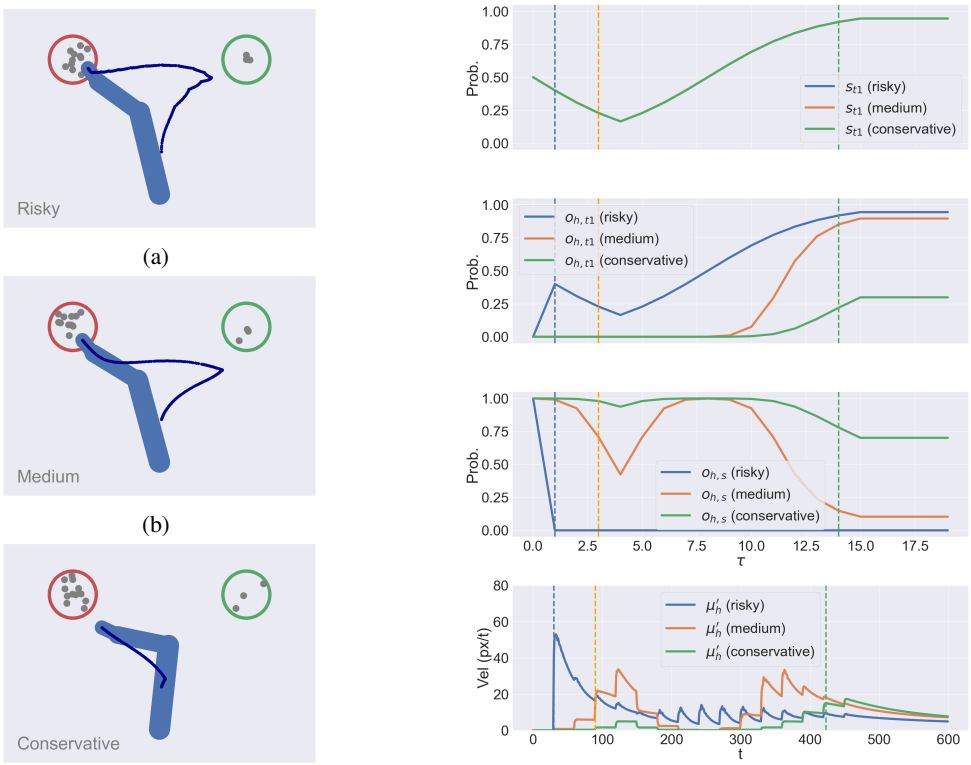


Figure 3: Motor planning with (a) a risky strategy (high urgency, or $\alpha_h = 0$), (b) a medium strategy (medium urgency, or $\alpha_h = 0.45$), and (c) a conservative strategy (low urgency, or $\alpha_h = 0.5$). (d) First panel: dynamics of the discrete hidden state of the first target over discrete time τ, which in this case are the same for every strategy. Second panel: dynamics of the discrete variable o_{t1} of the first target. Third panel: dynamics of the discrete variable o_s of staying in position. The fourth panel shows the L2-norm of the belief $\boldsymbol{\mu}'_h$ over the hand velocity in continuous time t. The vertical dashed lines represent the movement onset for each strategy (using $||\boldsymbol{\mu}'_h|| > 10$ as threshold).

where $\boldsymbol{\mu}_h$ is the belief over the hand position $\boldsymbol{x}_h$, λ is an attractor gain, while $\boldsymbol{p}_{t1}$ and $\boldsymbol{p}_{t2}$ are the positions of the two targets – assumed to be known and fixed. This mechanism implements the simultaneous preparation of competing motor plans, as also reported in monkey premotor cortex [40].

The mapping between discrete and continuous signals is done via Bayesian model average – as explained in Appendix B. In particular, a prior $\boldsymbol{\eta}'_{x,h}$ over the hand velocity $\boldsymbol{x}'_h$ is computed by weighting the three potential trajectories (to reach the two targets and to stay) with their respective probabilities $\boldsymbol{o}_h$ that the agent plans for a given discrete goal:

$$\boldsymbol{\eta}'_{x,h} = \boldsymbol{o}_h \cdot \boldsymbol{f}(\boldsymbol{\mu}_h) \tag{5}$$

This desired velocity enters the update of the continuous hidden states as a dynamics prediction error $\boldsymbol{\varepsilon}_{x,h} = \boldsymbol{\mu}'_h - \boldsymbol{\eta}'_{x,h}$, expressing a composite motion that the continuous model will realize. The inverse process, i.e., the inference of the current trajectory, is explained in the following section, and more details are found in Appendix B.

Figure 3 illustrates the effect on movement onset and velocity with three levels of urgency, during an incongruent trial. Since $k_h = 0$, the evidence accumulation is the same for the three cases, but the trajectories change depending on the agent's urgency to move, giving rise to risky, medium and conservative strategies. High and intermediate levels of urgency produce riskier strategies that initially move toward the wrong target and then manifest changes of mind. Low urgency produces a conservative strategy that moves directly toward the correct target, but has higher reaction times when the two target probabilities are too close, failing to complete the trial within the deadline. This is because the trajectories generated by the discrete model are constantly weighted by the *stay* dynamics,

hence high uncertainty α_h (which means low urgency) results not just in late movement onset but also in slower motion. This simulation illustrates that manipulating urgency provides flexibility in the link between evidence accumulation and movement dynamics. With high urgency, the agent moves earlier and takes the risk of failing, whereas with low urgency, the agent may wait until it accumulates sufficient evidence to reach very high confidence about the correct target. Interestingly, urgency and speed of evidence accumulation can interact, as shown in Figure S1.

2.3 The third pathway: motor inference and commitment

Various studies found that participants who initiate a movement show high commitment to the initially selected target – even in the face of contrasting evidence – when the costs required to reach the alternative target increase [26, 25]. Interestingly, in our model, this commitment emerges automatically during model inversion. As the agent jointly infers the correct target and the optimal discrete hand dynamics to reach it, there is a reciprocal interaction between the (top-down) pathway of *motor planning* and its dual (bottom-up) pathway of *motor inference* (Figure 1b). This is because in our model, not only the agent makes predictions over the cue that will be observed next (i.e., $\boldsymbol{A}_c\boldsymbol{s}$), but also over the hand trajectory (i.e., $\boldsymbol{A}_h\boldsymbol{s}$). The latter prediction entails a causal relation between discrete goals (in this case, the probability of the two targets) and the agent's movements (here, the dynamics needed to reach the targets) – as explained in Appendix B. The key implication of this perspective is that the probability of the two targets can be estimated from the hand trajectory itself ($\boldsymbol{o}_h$). As a consequence, a *self-evidencing* mechanism takes place, during which a target inferred by some cues produces a plan to reach it, which in turn confirms the initial agent's estimate. In other words, via motor inference, movement *stabilizes* (i.e., lowers the uncertainty about) the decision and creates commitment to the initially selected target.

The rich interplay between evidence accumulation, motor planning and motor inference gives rise to three competitive processes. The first competition occurs when estimating the continuous hidden states of the hand trajectory, defined by the generalized beliefs $\dot{\tilde{\boldsymbol{\mu}}}_h = [\boldsymbol{\mu}_h, \boldsymbol{\mu}'_h]$:

$$\dot{\tilde{\boldsymbol{\mu}}}_h = \mathcal{D}\tilde{\boldsymbol{\mu}}_h - \partial_{\tilde{\mu}_h}\mathcal{F} = \begin{bmatrix} \boldsymbol{\mu}'_h - \boldsymbol{\pi}_{\eta,h}\boldsymbol{\varepsilon}_{\eta,h} + \partial_{\mu_h}\boldsymbol{g}_v^T\boldsymbol{\pi}_v\boldsymbol{\varepsilon}_v + \partial_{\mu_h}\boldsymbol{f}^T\boldsymbol{\pi}_{x,h}\boldsymbol{\varepsilon}_{x,h} \\ \partial_{\mu'_h}\boldsymbol{g}_v'^T\boldsymbol{\pi}'_v\boldsymbol{\varepsilon}'_v - \boldsymbol{\pi}_{x,h}\boldsymbol{\varepsilon}_{x,h} \end{bmatrix} \tag{6}$$

This update rule includes a likelihood term coming from visual observations $\boldsymbol{y}_v$ and $\boldsymbol{y}'_v$ for both orders, which keeps the belief close to the actual hand trajectory; the dynamics prediction error $\boldsymbol{\varepsilon}_{x,h}$ defined in the previous section, affecting both orders either as a forward or backward message; and a prior prediction error $\boldsymbol{\varepsilon}_{\eta,h}$ that biases the belief over the hand position. The latter is linked, through forward kinematics, to an intrinsic continuous model encoding proprioceptive trajectories (e.g., expressed in joint angles). As a consequence, the backward message sent from the belief over the hand position to the intrinsic model performs inverse kinematics, eventually driving action. See Appendix C and [42, 43] for more details about kinematic inference.

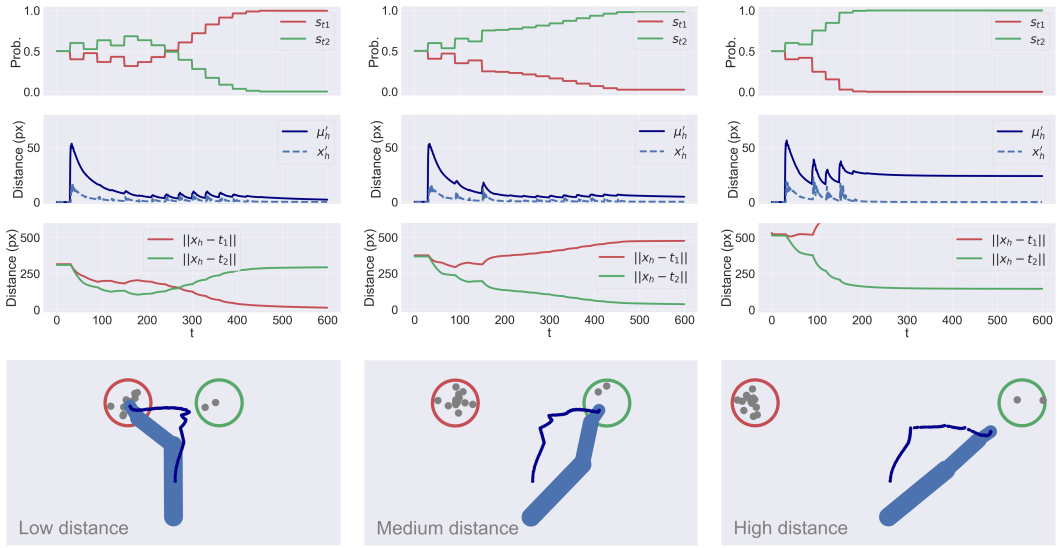
The second (and most interesting) competition happens when this continuous belief clashes with the desired hand dynamics needed to reach the selected target. While from a top-down perspective the dynamics defined in Equation 4 act as potential trajectories averaged by the probabilities $\boldsymbol{o}_h$, from a bottom-up perspective they are used to infer the most likely explanation of the real trajectory. More formally, the discrete hand dynamics $\boldsymbol{o}_h$ at time τ is found via Bayesian model comparison, i.e., by comparing the discrete prediction $\boldsymbol{A}_h\boldsymbol{s}_\tau$ with the continuous evidence $\boldsymbol{\mathcal{L}}_h$ of the hand trajectory, accumulated over a time window T:

$$\boldsymbol{o}_{h,\tau} = \sigma_w(\ln \boldsymbol{A}_h\boldsymbol{s}_\tau + \int_0^T \boldsymbol{\mathcal{L}}_h dt) \tag{7}$$

As before, σ_w is a weighted softmax whose slope w here controls how fast the transition between different dynamics occurs. High and low values of σ_w correspond to abrupt and gradual movement onsets, respectively. For each mth dynamics of Equation 4, the log evidence $\mathcal{L}_{h,m}$ is computed by:

$$\mathcal{L}_{h,m} = \frac{1}{2}(\boldsymbol{\mu}_{h,m}'^T\boldsymbol{p}_{x,h}\boldsymbol{\mu}'_{h,m} - \boldsymbol{f}_m(\boldsymbol{\mu}_h)^T\boldsymbol{\pi}_{x,h}\boldsymbol{f}_m(\boldsymbol{\mu}_) - \boldsymbol{\mu}_h'^T\boldsymbol{p}_{x,h}\boldsymbol{\mu}'_h + \boldsymbol{\eta}_{x,h}'^T\boldsymbol{\pi}_{x,h}\boldsymbol{\eta}'_{x,h}) \tag{8}$$

More details about the reduced posteriors $\boldsymbol{\mu}'_{h,m}$ and prior precisions $\boldsymbol{p}_{x,h}$ can be found in Appendix B, and in [44, 35, 36]. Here, we note that each potential dynamics function $\boldsymbol{f}_m$ is compared to the



(a) (b) (c)

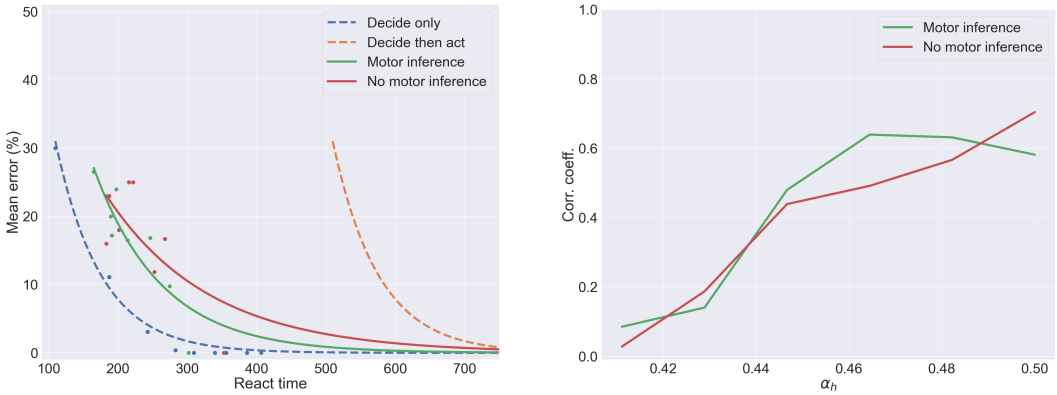
Figure 4: Commitment during an incongruent trial, with (a) low, (b) medium, and (c) high distance between the two targets, with $k_h = 0.15$ and $\alpha_h = 0$. For each condition, the first plot shows the discrete hidden states $\boldsymbol{s}$; the second plot shows the L2-norms of the estimated and true hand velocities $\boldsymbol{\mu}'_e$ and $\boldsymbol{x}'_e$; the third plot shows the distances between the hand and the two targets. For each condition, the second row also shows the agent's final trajectories, in dark blue.

current dynamics $\boldsymbol{\mu}_h$ inferred via sensory observations. As a result, the log evidence $\boldsymbol{\mathcal{L}}_h$ assigns higher values to the potential trajectories that better match the real one.

The third and final competition emerges at the intersection between motor inference and evidence accumulation, as highlighted in Equation 2. Since the discrete hand dynamics are constantly steered toward the current hand trajectory, the latter becomes a predictor for the correct target. This ultimately produces a commitment toward the chosen dynamics – expressed in the rightmost term of Equation 2.

One way to appreciate commitment is considering that as the distance between the two targets increases human participants make fewer changes of mind [25]. In keeping, when simulating 100 *neutral* trials with targets at different distances, we found that changes of mind were more frequent with targets at low distance ($n = 29$) than at medium ($n = 20$) and high distance ($n = 8$) – see Figure 4 for some sample trials. The reason for this result lies in Equation 7. Since the discrete hand dynamics $\boldsymbol{o}_h$ are used to infer the correct choice alternative, closer targets are scored with greater probabilities than farther targets. Furthermore, since the softmax function amplifies the differences between current and potential trajectories, a target is assigned an increasingly lower probability, the farther it is from the other one. This is evident in Figure S2, which shows the potential trajectories to reach both targets for the entire trial, the actual trajectory, and the log evidence accumulated over time. Another way to appreciate commitment is by comparing the agent's behavior during *congruent* and *incongruent* conditions, with motor inference ($k_h > 0$) and without it ($k_h = 0$) – see Figure S3. When motor inference is active in the *congruent* condition (Figure S3b), the hand dynamics stabilize the decision: the target is inferred faster than without motor inference (Figure S3a) and the second wrong cue (observed at $\tau = 9$) is ignored. When motor inference is active in the *incongruent* condition (Figure S3d), the agent could commit to a wrong decision ignoring contrasting evidence – a behavior that is not observed without motor inference (Figure S3c).

Finally, we considered the tradeoffs between various models: two serial models, one having instantaneous decision (*decide-only model*) and another with fixed decision time of 480 time steps (*decide-then-act model*); and two embodied choice models, one with motor inference ($k_h = 0.15$) and one without it ($k_h = 0$). For this comparison, we simulated 100 *neutral* trials, by varying the urgency α_h and the drift rate α_c to obtain a wide range of solutions. Figure 5a shows that the speed-accuracy tradeoff for the two embodied choice models is better than the decide-then-act model, as they are closer to the decide-only (ideal) model that moves instantaneously. Furthermore, Figure 5b shows that



(a) (b)

Figure 5: (a) Speed-accuracy curves for four models: two serial models (decide-only, decide-then-act) and two active inference models (with and without motor inference). The former two were sampled by varying the decision threshold in 500 trials, while the latter two were sampled by running 100 trials for different values of urgency (from medium-high urgency, i.e., $\alpha_h = 0.3$, to low urgency, i.e., $\alpha_h = 0.5$) and drift rate (from medium uncertainty, i.e., $\alpha_c = 0.4$, to high uncertainty, i.e., $\alpha_c = 0.49$). The samples were then fitted into a curve. Motor inference was realized with $k_h = 0.15$. For all conditions, $\alpha_c = 0.4$. To allow the agents to complete the trials on time with low levels of urgency, we set trial duration to 750 time steps. (b) Pearson product-moment correlation coefficient between the belief over the hand trajectory $\boldsymbol{\mu}'_h$ and the probability of the correct choice, computed for two conditions (with and without motor inference), with 500 neutral trials per condition, and with different levels of urgency.

across various levels of urgency α_h, there is a significant correlation between the speed of embodied choice levels and confidence (i.e., probability of the correct choice) as reported empirically [45].

2.4 The fourth pathway: statistical learning and habit formation

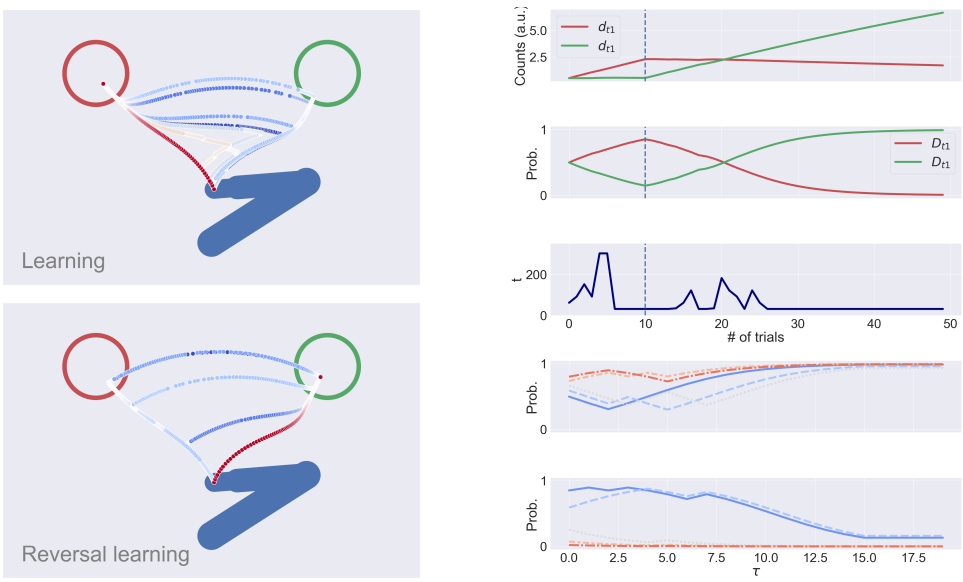
During various cognitive tasks such as the Flanker [46] and the Posner task [47], it is possible to learn statistical regularities, such as the probability of the correct response or the validity of cues across trials. In these tasks, trial sequence effects are often reported, indicating that participants form expectations across trials that influence their subsequent responses and movements [48]. The fourth pathway of our model implements this kind of statistical learning, which simply amounts to keeping count of the Dirichlet priors over the discrete hidden states $\boldsymbol{s}$ across trials. After every trial, these counts are updated according to:

$$\boldsymbol{d}_n = \omega \boldsymbol{d}_{n-1} + \eta \boldsymbol{s} \tag{9}$$

where n is the trial number, ω is a forgetting factor of older trials, and η is the learning rate of new trials (usually initialized with a reasonably high value, reflecting a high confidence over the prior belief, $\boldsymbol{d}_0$). Then, the counts are passed through a softmax function to compute the priors of the correct response for the next trial:

$$\boldsymbol{D}_n = \sigma(\boldsymbol{d}_n) \tag{10}$$

Figure 6 shows the effects of learning the prior over the correct response, during 50 incongruent trials. During the first 10 trials, the correct (left) response remains stable and then it is reversed. During the early trials of the first (learning) phase (dark blue in Figure 6a), the agent moves toward the wrong direction and then changes mind. However, in later trials (dark red in Figure 6a) it gradually begins to move early toward the correct target, anticipating the transition in the accumulation of wrong cues. In parallel, movement onset decreases (Figure 6b). These results show that a strong prior can overcome conflicting evidence. After the reversal at trial 10, the discrete prior for the first target slowly decreases, as the Dirichlet counts for the second target begin to increase (Figure 6b). In early trials, movement curvature increases and movement onset is slower, as the agent is uncertain about



(a) (b)

Figure 6: Statistical learning of the prior for the correct choice over 50 incongruent trials. The correct choice is fixed in the first 10 trials, but is reversed in the next 40 trials. (a) Hand trajectories in equally spaced trials, during learning (top) and reversal learning (bottom). Dark blue trajectories represent early trials, while dark red trajectories are late trials. Here, $k_e = 0$, $\omega = 0.99$, $\eta = 0.2$, $\alpha_c = 0.4$, and $\alpha_h = 0.4$. (b) The five panels show Dirichlet counts $\boldsymbol{d}$; discrete priors $\boldsymbol{D}$; time step of movement onset across trials; discrete hidden state s_1 in 5 equally spaced trials for learning; and for reversal learning over discrete time τ. The vertical dashed lines indicate the time step when reversal occurs.

the correct distribution underneath the cue sampling. In late trials, movement curvature decreases and movement onset fastens, as the agent learns the novel contingencies.

These results show that our model can incorporate sequence effects that emerge during cognitive tasks [48]. Note that while we focus on learning the prior probability of the correct choice, other model parameters, such as the uncertainties of the likelihood matrices $\boldsymbol{A}_c$ and $\boldsymbol{A}_h$, could be updated using the same approach.

3 Discussion

For many years, the dominant view regarding human and animal behaviors has been that of a serial, *decide-then-act* strategy. However, various studies show that during embodied decisions that require simultaneously specifying and selecting between alternative action plans, the serial view is insufficient. These studies report early movement onset, changes of mind, and the influence of motor costs over decisions, suggesting that decision and action processes unfold in parallel and reciprocally influence each other [27, 21, 12, 13, 10]. Here, we show that these signatures of embodied decisions emerge naturally in active inference: a framework that jointly optimizes decisions and actions under a free energy minimization imperative [31, 29]. Our simulations highlight that four model pathways – evidence accumulation, motor planning, motor inference, and statistical learning – form a closed loop, allowing decision and action processes to influence each other reciprocally. The resulting embodied models attain a better speed-accuracy tradeoff, compared to serial models – suggesting that they could confer ecological advantages (Figure 5a).

An innovative aspect of our model is the reciprocal interaction between motor planning and motor inference. During motor planning, the agent's inference of the correct choice generates predictions about the next discrete hand dynamics, which are converted into a continuous motor plan to reach the associated target. In turn, during motor inference, the agent uses the action dynamics as evidence for the correct choice. This mechanism implies that movement stabilizes decisions and creates commitment. Furthermore, it explains key aspects of embodied decisions, such as the fact that motor

costs apparent before a task [22] or changing during it [26] influence decision outcomes. Note that our model does not explicitly compute motor costs, but rather the probabilities of discrete hand dynamics $\boldsymbol{o}_h$, which could be equivalently considered as such following the duality of inferential formulations (that use probabilities) and control formulations (that use costs) [49, 50]. In short, the agent just tries to infer the correct target from the information at its disposal (including its own movements), and a high motor cost only means a potential dynamics that poorly explains the present context. Nonetheless, additional motor costs could be included in the model as prior preferences for biomechanically simpler movements [22].

Another key aspect of our model is the fact that by modulating the urgency to move, it can model a range of strategies – from riskier to more conservative – observed in empirical studies [41]. Future work might explore how the generative model advanced here could be inverted, to identify personalized parameters (e.g., a person's urgency) from behavioral data. While we covered several important aspects of embodied decisions in active inference, we kept the focus on the relationship between discrete decisions and continuous dynamics. A more realistic model would also consider discrete dynamics, which may be crucial for explaining how humans optimize the number of successes accumulated in a limited period – as analyzed in [33]. Furthermore, while our model explains how movement can stabilize decisions, it does not include other stabilization mechanisms, such as sensory gain modulation [51, 52, 53], which might be covered in future studies.

Acknowledgments

This research received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 952215 (TAILOR); the European Research Council under the Grant Agreement No. 820213 (ThinkAhead), the Italian National Recovery and Resilience Plan (NRRP), M4C2, funded by the European Union – NextGenerationEU (Project IR0000011, CUP B51E22000150006, "EBRAINS-Italy"; Project PE0000013, "FAIR"; Project PE0000006, "MNESYS"), the PRIN PNRR P20224FESY, and the European Union's Horizon H2020-EIC-FETPROACT-2019 Programme for Research and Innovation under Grant Agreement 951910. The GEFORCE Quadro RTX6000 and Titan GPU cards used for this research were donated by the NVIDIA Corporation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Roger Ratcliff and Gail McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922, April 2008.

[2] Marius Usher and James L McClelland. The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, 108(3):550, 2001.

[3] Michael N Shadlen and William T Newsome. Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *Journal of neurophysiology*, 86(4):1916–1936, 2001.

[4] Paul Cisek and Alexandre Pastor-Bernier. On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655):20130479, 2014.

[5] Giovanni Pezzulo and Paul Cisek. Navigating the Affordance Landscape: Feedback Control as a Process Model of Behavior and Cognition. *Trends in Cognitive Sciences*, 20(6):414–424, 2016.

[6] Jeremy Gordon, Antonella Maselli, Gian Luca Lancia, Thomas Thiery, Paul Cisek, and Giovanni Pezzulo. The road towards understanding embodied decisions. *Neuroscience & Biobehavioral Reviews*, 131:722–736, 2021.

[7] Antonella Maselli, Jeremy Gordon, Mattia Eluchans, Gian Luca Lancia, Thomas Thiery, Riccardo Moretti, Paul Cisek, and Giovanni Pezzulo. Beyond simple laboratory studies: developing sophisticated models to study rich behavior. *Physics of Life Reviews*, 2023.

[8] Arbora Resulaj, Roozbeh Kiani, Daniel M Wolpert, and Michael N Shadlen. Changes of mind in decision-making. *Nature*, 461(7261):263–266, 2009.

[9] Michael J Spivey, Marc Grosjean, and Günther Knoblich. Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102(29):10393–10398, 2005.

[10] Nathan J Wispinski, Jason P Gallivan, and Craig S Chapman. Models, movements, and minds: bridging the gap between decision making and action. *Annals of the New York Academy of Sciences*, 1464(1):30–51, 2020.

[11] Paul Cisek. Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485):1585–1599, 2007.

[12] Jonathan B Freeman, Rick Dale, and Thomas A Farmer. Hand in motion reveals mind in motion. *Frontiers in psychology*, 2:59, 2011.

[13] Joo-Hyun Song and Ken Nakayama. Hidden cognitive states revealed in choice reaching tasks. *Trends in cognitive sciences*, 13(8):360–366, 2009.

[14] Julien Michalski, Andrea M Green, and Paul Cisek. Reaching decisions during ongoing movements. *Journal of Neurophysiology*, 123(3):1090–1102, 2020.

[15] Gary A. Kane, Ryan A. Senne, and Benjamin B. Scott. Rat movements reflect internal decision dynamics in an evidence accumulation task. September 2023.

[16] Kaleb T. Kinder, Aaron T. Buss, and A. Caglar Tas. Tracking flanker task dynamics: Evidence for continuous attentional selectivity. *Journal of Experimental Psychology: Human Perception and Performance*, 48(7):771–781, July 2022.

[17] Laura Barca and Giovanni Pezzulo. Unfolding visual lexical decision in time. *PloS one*, 7(4):e35932, 2012.

[18] Laura Barca and Giovanni Pezzulo. Tracking second thoughts: Continuous and discrete revision processes during visual lexical decision. *PLoS One*, 10(2):e0116193, 2015.

[19] Michael Spivey. *The continuity of mind*. Oxford University Press, 2008.

[20] Charles W Eriksen and Derek W Schultz. Information processing in visual search: A continuous flow conception and experimental results. *Perception & psychophysics*, 25(4):249–263, 1979.

[21] Nathan F. Lepora and Giovanni Pezzulo. Embodied choice: How action influences perceptual decision making. *PLOS Computational Biology*, 11(4):e1004110, April 2015.

[22] Encarni Marcos, Ignasi Cos, Benoît Girard, and Paul FMJ Verschure. Motor cost influences perceptual decisions. *PLoS One*, 10(12):e0144841, 2015.

[23] Nobuhiro Hagura, Patrick Haggard, and Jörn Diedrichsen. Perceptual decisions are biased by the cost to act. *Elife*, 6:e18422, 2017.

[24] Eric Grießbach, Philipp Raßbach, Oliver Herbort, and Rouwen Cañal-Bruland. Embodied decisions during walking. *Journal of Neurophysiology*, 128(5):1207–1223, 2022.

[25] Diana Burk, James N. Ingram, David W. Franklin, Michael N. Shadlen, and Daniel M. Wolpert. Motor effort alters changes of mind in sensorimotor decision making. *PLoS ONE*, 9(3):e92681, March 2014.

[26] Ignasi Cos, Giovanni Pezzulo, and Paul Cisek. Changes of mind after movement onset depend on the state of the motor system. *Eneuro*, 8(6), 2021.

[27] Vassilios Christopoulos and Paul R Schrater. Dynamic integration of value information into a common probability currency as a theory for flexible decision making. *PLoS computational biology*, 11(9):e1004402, 2015.

[28] Paul Cisek. Making decisions through a distributed consensus. *Current opinion in neurobiology*, 22(6):927–936, 2012.

[29] 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.

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

[31] Thomas Parr, Giovanni Pezzulo, and Karl J Friston. Active inference: the free energy principle in mind, brain, and behavior. 2022.

[32] Matteo Priorelli, Federico Maggiore, Antonella Maselli, Francesco Donnarumma, Domenico Maisto, Francesco Mannella, Ivilin Peev Stoianov, and Giovanni Pezzulo. Modeling motor control in continuous-time Active Inference: a survey. *IEEE Transactions on Cognitive and Developmental Systems*, pages 1–15, 2023.

[33] Paul Cisek, Geneviève Aude Puskas, and Stephany El-Murr. Decisions in changing conditions: The urgency-gating model. *Journal of Neuroscience*, 29(37):11560–11571, 2009.

[34] Barbara A. Eriksen and Charles W. Eriksen. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics*, 16(1):143–149, January 1974.

[35] Matteo Priorelli and Ivilin Peev Stoianov. Hierarchical hybrid modeling for flexible tool use. 2024.

[36] Matteo Priorelli and Ivilin Peev Stoianov. Dynamic planning in hierarchical active inference. 2024.

[37] Jan Calalo, Truc Ngo, Seth Sullivan, Adam Roth, Rakshith Lokesh, John Buggeln, Kathryn Strand, Michael Carter, Isaac Kurtzer, and Joshua Cashaback. Reaching reflects ongoing deliberation prior to a decision. *Journal of Exercise, Movement, and Sport (SCAPPS refereed abstracts repository)*, 54(1), 2023.

[38] Christopher D Erb, Katie A Smith, and Jeff Moher. Tracking continuities in the flanker task: From continuous flow to movement trajectories. *Attention, Perception, & Psychophysics*, 83:731–747, 2021.

[39] Gabriele Gratton, Michael GH Coles, Erik J Sirevaag, Charles W Eriksen, and Emanuel Donchin. Pre-and poststimulus activation of response channels: a psychophysiological analysis. *Journal of Experimental Psychology: Human perception and performance*, 14(3):331, 1988.

[40] Paul Cisek and John F Kalaska. Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, 45(5):801–814, 2005.

[41] Cinzia Calluso, Giorgia Committeri, Giovanni Pezzulo, Nathan Lepora, and Annalisa Tosoni. Analysis of hand kinematics reveals inter-individual differences in intertemporal decision dynamics. *Experimental brain research*, 233:3597–3611, 2015.

[42] Matteo Priorelli, Giovanni Pezzulo, and Ivilin Peev Stoianov. Deep kinematic inference affords efficient and scalable control of bodily movements. *Proceedings of the National Academy of Sciences of the United States of America*, 120, 2023.

[43] M. Priorelli, G. Pezzulo, and I.P. Stoianov. Active vision in binocular depth estimation: A top-down perspective. *Biomimetics*, 8(5), 2023.

[44] M. Priorelli and I.P. Stoianov. Dynamic inference by model reduction. *bioRxiv*, 2023.

[45] Dror Dotan, Florent Meyniel, and Stanislas Dehaene. On-line confidence monitoring during decision making. *Cognition*, 171:112–121, 2018.

[46] Gabriele Gratton, Michael GH Coles, and Emanuel Donchin. Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121(4):480, 1992.

[47] Carlos M Gómez, Antonio Arjona, Francesco Donnarumma, Domenico Maisto, Elena I Rodríguez-Martínez, and Giovanni Pezzulo. Tracking the time course of bayesian inference with event-related potentials: A study using the central cue posner paradigm. *Frontiers in Psychology*, 10:1424, 2019.

[48] Wenting Ye and Markus F Damian. Effects of conflict in cognitive control: Evidence from mouse tracking. *Quarterly Journal of Experimental Psychology*, 76(1):54–69, 2023.

[49] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960.

[50] Emanuel Todorov. General duality between optimal control and estimation. In *2008 47th IEEE conference on decision and control*, pages 4286–4292. IEEE, 2008.

[51] Christopher L Buckley and Taro Toyoizumi. A theory of how active behavior stabilises neural activity: Neural gain modulation by closed-loop environmental feedback. *PLoS computational biology*, 14(1):e1005926, 2018.

[52] Kenneth D Harris and Alexander Thiele. Cortical state and attention. *Nature reviews neuroscience*, 12(9):509–523, 2011.

[53] Emilio Salinas and Terrence J Sejnowski. Book review: gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. *The Neuroscientist*, 7(5):430–440, 2001.

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

[55] 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.

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

[57] 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.

[58] Thomas Parr and Karl J. Friston. The Discrete and Continuous Brain: From Decisions to Movement—And Back Again Thomas. *Neural Computation*, 30:2319–2347, 2018.

[59] A. Tschantz, L. Barca, D. Maisto, C. L. Buckley, A. K. Seth, and G. Pezzulo. Simulating homeostatic, allostatic and goal-directed forms of interoceptive control using active inference. *Biological Psychology*, 169:108266, 2022.

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

[61] 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.

[62] 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.

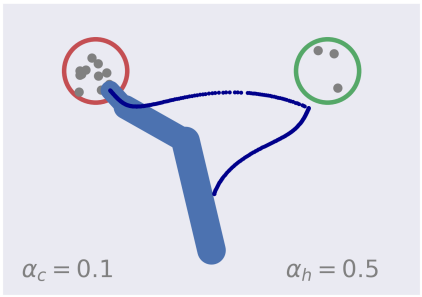
[63] Matthew Botvinick and Marc Toussaint. Planning as inference. *Trends in Cognitive Sciences*, 16(10):485–488, 2012.

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

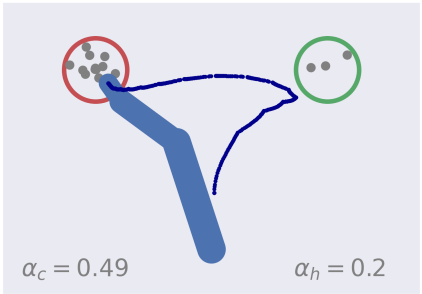
[65] Rick A. Adams, Stewart Shipp, and Karl J. Friston. Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, 218(3):611–643, 2013.

[66] Harriet Brown, Karl Friston, and Sven Bestmann. Active inference, attention, and motor preparation. *Frontiers in Psychology*, 2(SEP):1–10, 2011.

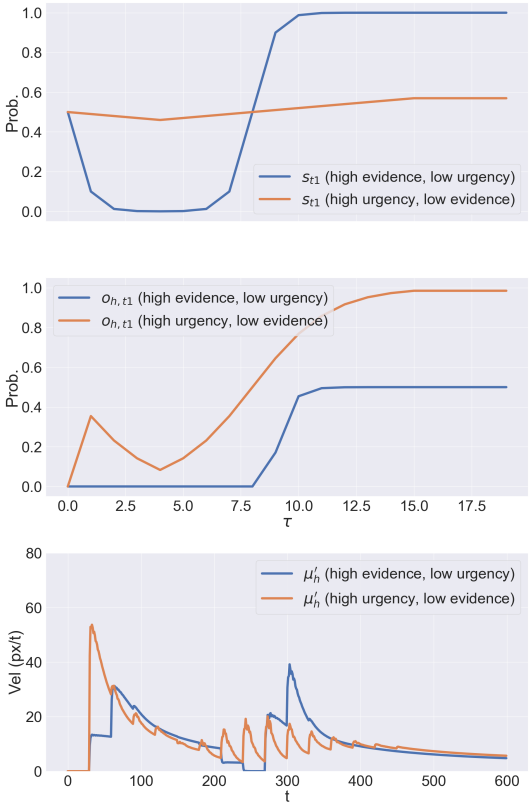
A Supplementary Figures



(a)



(b)



(c)

Figure S1: Interaction between urgency and speed of evidence accumulation. (a) Fast evidence accumulation ($\alpha_c = 0.1$) and low urgency to move ($\alpha_h = 0.5$). (b) Slow evidence accumulation ($\alpha_c = 0.49$) and high urgency to move ($\alpha_h = 0.2$). (c) While movement dynamics look similar, the evolution of the hidden states and hand dynamics are different.

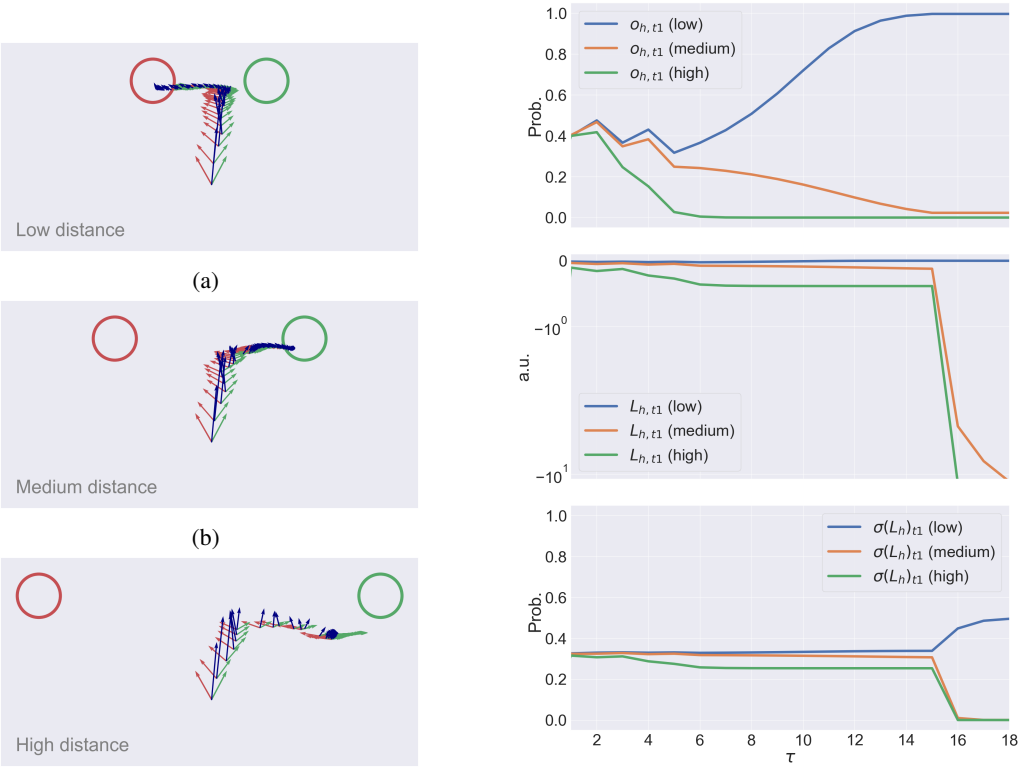


Figure S2: Commitment during an incongruent trial, with (a) low, (b) medium, and (c) high distance between the two targets; continued from Figure 4. (a-c) The panels show the direction and magnitude of the estimated velocity $\boldsymbol{\mu}'_h$ (in dark blue), the potential trajectory needed to reach the first target $\boldsymbol{f}_{t1}$ (in red) and the second target $\boldsymbol{f}_{t2}$ (in green), for each condition. These potential trajectories are used to estimate which of the two targets is more likely to have generated the current trajectory. (d) The top panel shows the discrete hand dynamics $o_{h,t1}$ of the first target over discrete time τ; the middle panel shows the log evidence $\mathcal{L}_{h,t1}$ of the hand trajectory associated with the first target, on a logarithmic scale; the bottom panel shows the normalized log evidence of the first target.

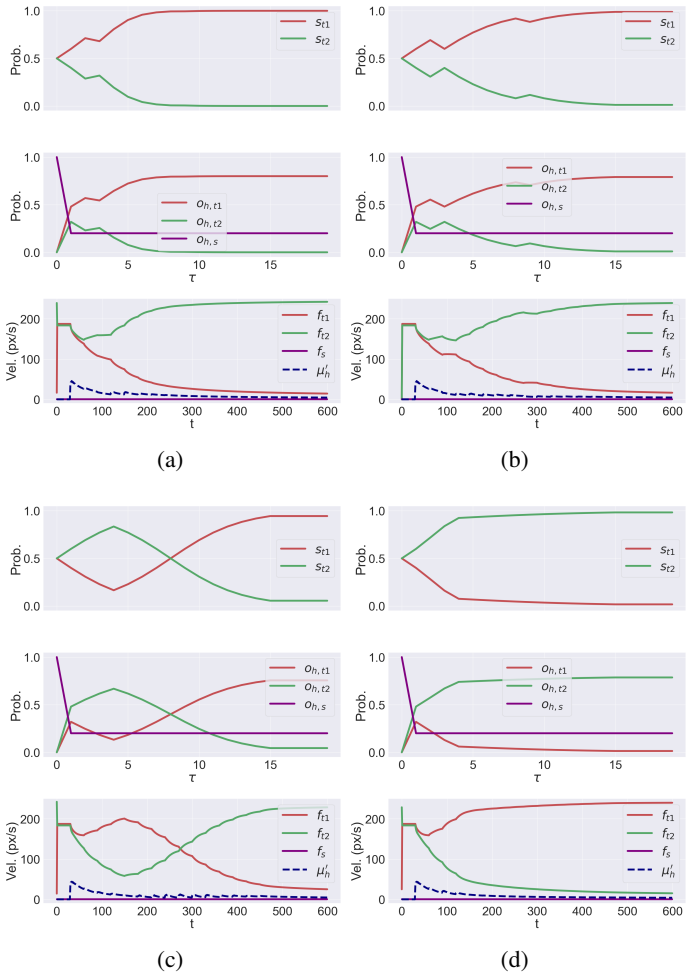


Figure S3: Commitment to an initially selected target resulting from motor inference. Panels a-b compare two agents, one without motor inference (a, $k_h = 0$) and one with it (b, $k_h = 0.2$), during a *congruent* trial. Panels c-d compare the same two agents, one without motor inference (a, $k_h = 0$) and one with it (b, $k_h = 0.2$), during an *incongruent* trial. In all conditions, $\alpha_h = 0.2$. The first (top) plot of each panel shows the discrete hidden states $\boldsymbol{s}$. The second plot shows the discrete hand dynamics for reaching both targets and staying in position. The third plot shows the L2-norms of the estimated hand velocity $\boldsymbol{\mu}'_h$, compared with the three potential trajectories $\boldsymbol{f}_{t1}$, $\boldsymbol{f}_{t2}$, and $\boldsymbol{f}_s$. Note that although the *stay* dynamics function $\boldsymbol{f}_s$ is (initially) the closest to the actual trajectory, the related probability $o_{h,s}$ decreases rapidly, as soon as the predictions $\boldsymbol{A}_h\boldsymbol{s}$ shift toward one of the targets, showing the top-down influence from choice to movement. The fourth (bottom) plot shows the agent's movement trajectories, in dark blue.

B Dynamic reduced models in active inference

Active inference is a computational theory that proposes a unifying paradigm to understand cognitive processing and behavior in living organisms. It is based upon the so-called *free energy principle* which states that, in order to survive, every creature must actively minimize surprise [54, 30, 55, 29].

Active inference models have been formulated both in discrete time and in continuous time. However, it is possible to argue that a comprehensive account of living organisms might require *hybrid* models, which combine both discrete and continuous time formulations [56, 57]. For example, in the human nervous system, the cerebral cortex might operate in a discrete state-space, while the low-level sensorimotor loops might be better understood in terms of continuous representations; the interface between the two is attributed to subcortical structures such as the thalamus or the superior colliculus [58]. Hybrid models have been used to simulate many scenarios such as pictographic reading [56], movements under neurological disorders [54], or interoceptive control [59]. Here, we briefly present a particular instance of such models useful to infer and act upon dynamic trajectories [44, 35, 36]. See [54] for more details about active inference and the free energy principle, and[60] about Bayesian model selection.

Discrete models share resemblances with Hidden Markov Models (HMMs) and are defined as partially observable Markov decision processes (POMDPs) [61, 62]. In particular, they are related to a subfield of machine learning known as *planning as inference* [63, 64]. Discrete models assume that organisms perceive the environment by optimizing an internal generative model, inferring how external causes lead to sensory signals within the environment (called generative process). Denoting by $\boldsymbol{s}$ the discrete hidden states, by $\boldsymbol{o}$ the discrete outcomes, by $\boldsymbol{\pi}$ the policies (which in active inference are sequences of actions), the agent's generative model after a discrete period T is factorized as:

$$p(\boldsymbol{s}_{1:T}, \boldsymbol{o}_{1:T}, \boldsymbol{\pi}) = p(\boldsymbol{s}_1) \cdot p(\boldsymbol{\pi}) \cdot \prod_{\tau=1}^{T} p(\boldsymbol{o}_\tau|\boldsymbol{s}_\tau) \cdot \prod_{\tau=2}^{T} p(\boldsymbol{s}_\tau|\boldsymbol{s}_{\tau-1}, \boldsymbol{\pi}) \tag{11}$$

where every distribution is assumed to be categorical:

$$\begin{aligned} p(\boldsymbol{s}_1) &= Cat(\boldsymbol{D}) & p(\boldsymbol{o}_\tau|\boldsymbol{s}_\tau) &= Cat(\boldsymbol{A}) \\ p(\boldsymbol{\pi}) &= Cat(\boldsymbol{E}) & p(\boldsymbol{s}_\tau|\boldsymbol{s}_{\tau-1}, \boldsymbol{\pi}) &= Cat(\boldsymbol{B}_{\pi,\tau}) \end{aligned} \tag{12}$$

Here, $\boldsymbol{D}$ is the prior over the initial state, $\boldsymbol{E}$ is the prior over policies, $\boldsymbol{A}$ is the likelihood (or observation) matrix, and $\boldsymbol{B}_{\pi,\tau}$ is the transition matrix. In order to find the causes of their perceptions, organisms try to infer the posterior distribution:

$$p(\boldsymbol{s}_{1:T}, \boldsymbol{\pi}|\boldsymbol{o}_{1:T}) = \frac{p(\boldsymbol{o}_{1:T}|\boldsymbol{s}_{1:T}, \boldsymbol{\pi})p(\boldsymbol{s}_{1:T}, \boldsymbol{\pi})}{p(\boldsymbol{o}_{1:T})} \tag{13}$$

However, this requires computing the intractable model evidence $p(\boldsymbol{o}_{1:T})$. Hence, organisms are supposed to implement some sort of approximate Bayesian inference by relying on an approximate posterior distribution $q(\boldsymbol{s}_{1:T}, \boldsymbol{\pi})$, and then minimizing the Kullback-Leibler (KL) divergence between the approximate and real posteriors:

$$\begin{aligned} 0 \leq &= D_{KL}[q(\boldsymbol{s}_{1:T}, \boldsymbol{\pi})||p(\boldsymbol{s}_{1:T}, \boldsymbol{\pi}|\boldsymbol{o}_{1:T})] \\ &= \mathop{\mathbb{E}}_{q}[\ln q(\boldsymbol{s}_{1:T}, \boldsymbol{\pi}) - \ln p(\boldsymbol{o}_{1:T}, \boldsymbol{s}_{1:T}, \boldsymbol{\pi})] + \ln p(\boldsymbol{o}_{1:T}) \\ &= \mathcal{F} + \ln p(\boldsymbol{o}_{1:T}) \end{aligned} \tag{14}$$

The KL divergence still requires the computation of the log evidence $p(\boldsymbol{o}_{1:T})$; however, we can exploit its non-negativity to instead minimize the first RHS term of the second line of Equation 14 – called variational free energy (VFE) and known in machine learning as the negative *ELBO* – which ensures that surprise $-\ln p(\boldsymbol{o}_{1:T})$ is minimized. Then, expressing the approximate posterior by its sufficient statistics $\boldsymbol{s}_{\pi,\tau}$ and conditioning upon a specific policy:

$$\begin{aligned} p(\boldsymbol{s}_{1:T}|\boldsymbol{o}_{1:T}, \boldsymbol{\pi}) \approx q(\boldsymbol{s}_{1:T}, \boldsymbol{\pi}) &= q(\boldsymbol{\pi}) \prod_{\tau}^{T} q(\boldsymbol{s}_\tau|\boldsymbol{\pi}) \\ q(\boldsymbol{\pi}) &= Cat(\boldsymbol{\pi}) \\ q(\boldsymbol{s}_\tau|\boldsymbol{\pi}) &= Cat(\boldsymbol{s}_{\pi,\tau}) \end{aligned} \tag{15}$$

We can infer the most likely discrete hidden state at time τ and under policy π by computing the gradient of the VFE $\mathcal{F}_\pi$ of that policy and applying a softmax function to ensure that it is a proper probability distribution:

$$\boldsymbol{s}_{\pi,\tau} = \sigma(\ln \boldsymbol{B}_{\pi,\tau-1}\boldsymbol{s}_{\pi,\tau-1} + \ln \boldsymbol{B}^T_{\pi,\tau}\boldsymbol{s}_{\pi,\tau+1} + \ln \boldsymbol{A}^T\boldsymbol{o}_\tau) \tag{16}$$

Via VFE minimization, organisms are able to capture the optimal representation of the environment; however, they are unable to perform any sort of future planning. To do this, they additionally consider unobserved outcomes as random variables, and they infer the most likely policy or sequence of actions that will lead to their preferred outcomes. More formally, if we denote by $p(\boldsymbol{o}_\tau|\boldsymbol{C})$ the probability distribution encoding the agent's preferred outcomes at some future point τ, the optimal policy is found by minimizing the free energy that the agent expects to perceive in the future – called *expected free energy* (EFE) and denoted by $\mathcal{G}$:

$$\begin{aligned}
\boldsymbol{\pi} &= \sigma(\ln \boldsymbol{E} - \mathcal{G}) \\
\mathcal{G}_\pi &\approx \sum_\tau D_{KL}[q(\boldsymbol{o}_\tau|\boldsymbol{\pi})||p(\boldsymbol{o}_\tau|\boldsymbol{C})] + \underset{q(\boldsymbol{s}_\tau|\boldsymbol{s}_{\tau-1},\boldsymbol{\pi})}{\mathbb{E}}[H[p(\boldsymbol{o}_\tau|\boldsymbol{s}_\tau)]] \\
&= \sum_\tau \boldsymbol{o}_{\pi,\tau}(\ln \boldsymbol{o}_{\pi,\tau} - \boldsymbol{C}_\tau) + \boldsymbol{s}_{\pi,\tau}\boldsymbol{H}
\end{aligned} \tag{17}$$

where:

$$\boldsymbol{o}_{\pi,\tau} = \boldsymbol{A}\boldsymbol{s}_{\pi,\tau} \qquad \boldsymbol{C}_\tau = \ln p(\boldsymbol{o}_\tau|\boldsymbol{C}) \qquad \boldsymbol{H} = -diag(\boldsymbol{A}^T \ln \boldsymbol{A}) \tag{18}$$

Equation 17 underlies a stark difference with theories of optimal control and reinforcement learning. These theories assume that hidden states have intrinsic values and that agents infer the optimal policy that maximizes the accumulation of future rewards obtained from the environment. Rather, active inference only assumes that each organism believes that the environment will evolve in a specific way, determined by its phenotype. Under this view, action is just another way, complementary to perception, to minimize free energy (hence, surprise), i.e., to reduce the difference between the agent's prior beliefs and the actual generative process. In other words, by acting, they make future observations coherent with their internal model – a process known as *self-evidencing*. Furthermore, the two components of the second and third lines of Equation 17 entail the well-known tradeoff between exploitation and exploration, with the addition that the latter, also called *ambiguity*, is involved in itinerant and novelty-seeking behavior.

To get a discrete model working with the richness of the continuous input, it is linked to an active inference continuous model [56, 57]. The latter is highly similar to the discrete model described above, with the difference that now only the VFE is minimized (for this reason, a continuous model alone is not capable of advanced decision-making). The standard way to link the two models is by letting a discrete outcome generate a *causal variable* (e.g., the position of a target to reach) that in turn produces a continuous trajectory; in this way, inverting the model means to infer the target based on the continuous trajectory, and then finding the discrete outcome that best explains the inferred target by comparing it with some fixed positions that the agent knows a-priori. In order to render the agent's representation more flexible, we can instead generate a continuous trajectory directly by a discrete outcome. Here, we briefly describe this alternative approach.

First, we model the continuous environment with the following non-linear stochastic equations:

$$\begin{aligned}
\tilde{\boldsymbol{y}} &= \tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}}) + \boldsymbol{w}_y \\
\mathcal{D}\tilde{\boldsymbol{x}} &= \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}) + \boldsymbol{w}_x
\end{aligned} \tag{19}$$

where $\tilde{\boldsymbol{x}}$ are continuous hidden states, $\tilde{\boldsymbol{y}}$ are continuous observations, $\tilde{\boldsymbol{g}}$ is a likelihood function defining how hidden states cause observations, $\tilde{\boldsymbol{f}}$ is a dynamics function expressing how the hidden states evolve, and the letter w indicates Gaussian noise terms. Note that the symbol $\sim$ denotes generalized coordinates of motion encoding instantaneous trajectories (e.g., position, velocity, acceleration, and so on), which in a continuous formulation replace the future states. Also, $\mathcal{D}$ is a differential operator that shifts every coordinate of motion by one. The joint distribution of this *hybrid* generative model is factorized into the following:

$$p(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \boldsymbol{o}_\tau) = p(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}})p(\tilde{\boldsymbol{x}}|\boldsymbol{o}_\tau)p(\boldsymbol{o}_\tau) \tag{20}$$

where the first two distributions are assumed to be Gaussian, while the last one is the categorical (likelihood) distribution defined before:

$$\begin{aligned} p(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}) &= \mathcal{N}(\tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}}), \tilde{\boldsymbol{\Sigma}}_y) \\ p(\tilde{\boldsymbol{x}}|\boldsymbol{o}_\tau) &= \mathcal{N}(\boldsymbol{\eta}, \tilde{\boldsymbol{\Sigma}}_x) \\ p(\boldsymbol{o}_\tau) &= Cat(\boldsymbol{A}) \end{aligned} \tag{21}$$

Here, $\boldsymbol{\eta}$ is the mean (also called *full prior*) of a complex model that defines the actual evolution of the hidden states. We suppose that the agent maintains M probability distributions which are reduced versions of this full model. Each reduced distribution corresponds to a particular *discrete dynamics* $o_{\tau,m}$ and encodes a specific way that the agent thinks the environment may evolve [35, 36]:

$$p(\tilde{\boldsymbol{x}}|o_{\tau,m}) = \mathcal{N}(\tilde{\boldsymbol{f}}_m(\tilde{\boldsymbol{x}}), \tilde{\boldsymbol{\Sigma}}_{x,m}) \tag{22}$$

where $\tilde{\boldsymbol{f}}_m$ is the dynamics function of the mth reduced model, and $\tilde{\boldsymbol{\Sigma}}_{x,m}$ the related precision. These alternative hypotheses act as empirical priors at the lower continuous level. Following the theory of Bayesian model reduction, in order to infer the posterior of the full model we first have to compute the posterior probability of each reduced model. In fact, reduced means that the likelihood of some data is equal to that of the full model and the only difference rests upon the specification of the priors – hence, the posterior of a reduced model can be expressed in terms of the posterior of the full model:

$$p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, o_{\tau,m}) = p(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{o}_\tau)\frac{p(\tilde{\boldsymbol{x}}|o_{\tau,m})p(\tilde{\boldsymbol{y}}|\boldsymbol{o}_\tau)}{p(\tilde{\boldsymbol{x}}|\boldsymbol{o}_\tau)p(\tilde{\boldsymbol{y}}|o_{\tau,m})} \tag{23}$$

As in the previous discrete formulation, we introduce a full and M approximate posteriors, here assumed to be Gaussian:

$$\begin{aligned} q(\tilde{\boldsymbol{x}}) &= \mathcal{N}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{C}}_x) \\ q(\tilde{\boldsymbol{x}}|m) &= \mathcal{N}(\tilde{\boldsymbol{\mu}}_m, \tilde{\boldsymbol{C}}_{x,m}) \end{aligned} \tag{24}$$

where $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}_m$ are the full and reduced posterior means (also called *beliefs*), while $\tilde{\boldsymbol{C}}_x$ and $\tilde{\boldsymbol{C}}_{x,m}$ are the full and reduced posterior precisions. Replacing the real posteriors with these approximate posteriors, we can write the free energy of each reduced model in terms of the full model. As before, maximizing each reduced free energy makes it approximate the log evidence:

$$\mathcal{F}(o_{\tau,m}) = \mathcal{F}(\boldsymbol{o}_\tau) - \ln \int \frac{p(\tilde{\boldsymbol{x}}|o_{\tau,m})}{p(\tilde{\boldsymbol{x}}|\boldsymbol{o}_\tau)} q(\tilde{\boldsymbol{x}}) d\tilde{\boldsymbol{x}} \approx \ln p(\tilde{\boldsymbol{y}}|o_{\tau,m}) \tag{25}$$

This relation implies that the free energy related to each discrete dynamics $o_{\tau,m}$ can be found from the approximate posterior $q(\tilde{\boldsymbol{x}})$ of the full model, without directly computing the reduced posteriors. With the Gaussian approximation of Equation 24, the mth reduced free energy breaks down to a simple formula:

$$\begin{aligned} \mathcal{F}(o_{\tau,m}) = &\frac{1}{2} \ln |\tilde{\boldsymbol{\Pi}}_{x,m} \tilde{\boldsymbol{P}}_x \tilde{\boldsymbol{C}}_{x,m} \tilde{\boldsymbol{\Sigma}}_x| \\ &- \frac{1}{2}(\tilde{\boldsymbol{\mu}}^T \tilde{\boldsymbol{P}}_x \tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}_m^T \tilde{\boldsymbol{P}}_{x,m} \tilde{\boldsymbol{\mu}}_m - \tilde{\boldsymbol{\eta}}^T \tilde{\boldsymbol{\Pi}}_x \tilde{\boldsymbol{\eta}} + \tilde{\boldsymbol{f}}_m(\tilde{\boldsymbol{\mu}})^T \tilde{\boldsymbol{\Pi}}_{x,m} \tilde{\boldsymbol{f}}_m(\tilde{\boldsymbol{\mu}})) \end{aligned} \tag{26}$$

where the mean and precision of mth reduced model are:

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_m &= \tilde{\boldsymbol{C}}_{x,m}(\tilde{\boldsymbol{P}}_x \tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\Pi}}_x \tilde{\boldsymbol{\eta}} + \tilde{\boldsymbol{\Pi}}_{x,m} \tilde{\boldsymbol{f}}_m(\tilde{\boldsymbol{\mu}})) \\ \tilde{\boldsymbol{P}}_{x,m} &= \tilde{\boldsymbol{P}}_x - \tilde{\boldsymbol{\Pi}}_x + \tilde{\boldsymbol{\Pi}}_{x,m} \end{aligned} \tag{27}$$

and we wrote the covariances $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{C}}$ in terms of precisions $\tilde{\boldsymbol{\Pi}}$ and $\tilde{\boldsymbol{P}}$. In short, the free energy assigns a score to each reduced model based on its fit with the full posterior, and two models i and j can be compared through a log-Bayes factor $\mathcal{F}(o_{\tau,i}) - \mathcal{F}(o_{\tau,j})$. Crucially, this posterior is a continuous trajectory and the agent's hypotheses are constantly updated via sensory observations. If a discrete step τ corresponds to a continuous period T, we integrate the reduced free energies over this period and compare them with the prior surprise $-\ln \boldsymbol{o}_\tau$. This results in an ascending message that infers the most likely discrete dynamics $\boldsymbol{o}_\tau$ that may have generated the current continuous observations:

$$\begin{aligned} E_{\tau,m} &= -\ln o_{\tau,m} - \int_0^T \mathcal{F}(o_{\tau,m}) dt \\ \boldsymbol{o}_\tau &= \sigma(-\boldsymbol{E}_\tau) \end{aligned} \tag{28}$$

Instead, the descending message is computed as a simple Bayesian model average of the agent's hypotheses, e.g., by weighting the discrete dynamics $o_{\tau,m}$ of a reduced model with the related dynamics function $\tilde{\boldsymbol{f}}_m(\tilde{\boldsymbol{\mu}})$:

$$\tilde{\boldsymbol{\eta}} = \sum_m o_{\tau,m} \tilde{\boldsymbol{f}}_m(\tilde{\boldsymbol{\mu}}) \tag{29}$$

Then, $\tilde{\boldsymbol{\eta}}$ embeds a prior over a continuous trajectory, which steers the inference of the continuous hidden states toward preferred outcomes. The posterior over the hidden states is finally found by computing the gradient of the free energy of the full model with respect to the mean $\tilde{\boldsymbol{\mu}}$ of the full approximate posterior, and updating the latter via gradient descent:

$$\dot{\tilde{\boldsymbol{\mu}}} - \mathcal{D}\tilde{\boldsymbol{\mu}} = -\partial_\mu \mathcal{F}(\boldsymbol{o}_\tau) = \partial \tilde{\boldsymbol{g}}^T \tilde{\boldsymbol{\Pi}}_y \tilde{\boldsymbol{\varepsilon}}_y + \partial_\mu \tilde{\boldsymbol{f}}^T \tilde{\boldsymbol{\Pi}}_x \tilde{\boldsymbol{\varepsilon}}_x - \mathcal{D}^T \tilde{\boldsymbol{\Pi}}_x \tilde{\boldsymbol{\varepsilon}}_x \tag{30}$$

This update is expressed in terms of message passing of the following prediction errors:

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}}_y &= \tilde{\boldsymbol{y}} - \tilde{\boldsymbol{g}}(\tilde{\boldsymbol{\mu}}) \\ \tilde{\boldsymbol{\varepsilon}}_x &= \mathcal{D}\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\eta}} \end{aligned} \tag{31}$$

and has a specular form of the discrete update of Equation 16, except that now the inference is done over a continuous path – hence the additional term $\mathcal{D}\tilde{\boldsymbol{\mu}}$. This minimization entails the process of perception, i.e., conforms the agent's beliefs to the perceived sensations. In order to conform the environment to the agent's beliefs (in short, to act), the free energy can be minimized with respect to the motor commands $\boldsymbol{a}$. This additional mechanism reduces to a minimization of sensory prediction errors:

$$\dot{\boldsymbol{a}} = -\partial_a \mathcal{F}(\boldsymbol{o}_\tau) = -\partial_a \tilde{\boldsymbol{g}}^T \tilde{\boldsymbol{\Pi}}_y \tilde{\boldsymbol{\varepsilon}}_y \tag{32}$$

where $\partial_a \tilde{\boldsymbol{g}}$ is an inverse mapping from sensations to actions. This mapping implements a dynamics inversion and is thought to be realized by classical reflex arcs in the spinal cord [65, 66].

C Inverse kinematics in active inference

Several methods have been proposed on how to realize inverse kinematics in active inference. The most common approach for simulating a reaching task is to encode a target location in extrinsic coordinates as a causal variable of a continuous model. This variable generates a prediction for the 1st-order (e.g., velocity) of the hidden states – encoding the agent's joint angles in intrinsic coordinates – which therefore act as a dynamic attractor toward the desired location. In this simple representation, the link between the causal variable and the hidden states performs an inverse kinematics of the target location, whose product (i.e., a possible agent's configuration with the hand at the target) is compared with the extrinsic position of the hand computed via forward kinematics. In this simple representation, both (intrinsic and extrinsic) reference frames are used in a single active inference level.

An alternative and more powerful approach exploits an aspect of the theory inherited from predictive coding, i.e., that the nervous system manages to approximate the real posterior by building a hierarchical architecture wherein a particular level acts as an observation for the level above and as a prior for the level below [42, 43]. In this way, higher levels can construct increasingly richer and more invariant representations of the environment, similar to deep generative models of neural networks. Since a level communicates only with the levels immediately above and below, the overall generative model can be factorized into independent distributions, and every level can be analyzed as in the general formulations of discrete and continuous active inference. Specifically, the hierarchical approach to inverse kinematics is to design a structure of two levels, wherein an *intrinsic unit* encoding the agent's joint angles has a causal influence on an *extrinsic unit* encoding the agent's hand. This approach follows the natural flow of the generative process, with an efficient decomposition between intrinsic and extrinsic dynamics. Here, we briefly describe this second approach.

The higher-level intrinsic unit $\mathcal{U}_i$ is governed by the following equations:

$$\begin{aligned} \tilde{\boldsymbol{y}}_p &= \tilde{\boldsymbol{g}}_p(\tilde{\boldsymbol{x}}_i) + \boldsymbol{w}_{y,p} \\ \mathcal{D}\tilde{\boldsymbol{x}}_i &= \tilde{\boldsymbol{f}}_i(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{v}}_i) + \boldsymbol{w}_{x,i} \end{aligned} \tag{33}$$

where $\tilde{\boldsymbol{y}}_p$ are generalized (e.g., position, velocity, and so on) proprioceptive observations, $\tilde{\boldsymbol{g}}_p$ are the generalized proprioceptive likelihoods, $\tilde{\boldsymbol{x}}_i$ are intrinsic hidden states encoding the instantaneous trajectory of the agent (e.g., every joint angle of the arm), $\tilde{\boldsymbol{v}}_i$ are intrinsic causal variables encoding the

trajectory priors, and $\tilde{\boldsymbol{f}}_i$ are the related generalized dynamics. Also, the letter w indicates Gaussian noises.

Instead, the lower-level extrinsic unit $\mathcal{U}_e$ follows the system:

$$\begin{aligned} \tilde{\boldsymbol{y}}_v &= \tilde{\boldsymbol{g}}_v(\tilde{\boldsymbol{x}}_e) + \boldsymbol{w}_{y,v} \\ \mathcal{D}\tilde{\boldsymbol{x}}_e &= \tilde{\boldsymbol{f}}_e(\tilde{\boldsymbol{x}}_e, \tilde{\boldsymbol{v}}_e) + \boldsymbol{w}_{x,e} \end{aligned} \tag{34}$$

where the letter v indicates the visual (or, more in general, exteroceptive) domain. The two units are related via connection between the 0th (i.e., position) temporal orders:

$$\boldsymbol{x}_e = \boldsymbol{g}_e(\boldsymbol{x}_i) = \boldsymbol{T}(\boldsymbol{x}_i) + \boldsymbol{w}_i \tag{35}$$

where $\boldsymbol{T}$ performs a forward kinematics between the agent's joint angles and its hand. As a consequence, the backward message from $\mathcal{U}_e$ to $\mathcal{U}_i$ realizes an inverse kinematics of an *extrinsic prediction error* conveying the difference between the prediction of $\mathcal{U}_i$ and the actual extrinsic belief of $\mathcal{U}_e$:

$$\partial \boldsymbol{g}_e^T \boldsymbol{\varepsilon}_e = \boldsymbol{J}^T(\boldsymbol{x}_e - \boldsymbol{T}(\boldsymbol{x}_i)) \tag{36}$$

In this way, extrinsic trajectories (e.g., linear or circular motions) are easily realized by defining appropriate dynamics in $\tilde{\boldsymbol{f}}_e$, which then travels back to infer the most likely agent's configuration corresponding to that trajectory. The updates of the belief over intrinsic and extrinsic hidden states are the following:

$$\begin{aligned} \dot{\tilde{\boldsymbol{\mu}}}_i - \mathcal{D}\tilde{\boldsymbol{\mu}}_i &= -\partial_{\mu_i}\mathcal{F} = \partial\tilde{\boldsymbol{g}}_e^T\tilde{\boldsymbol{\pi}}_e\tilde{\boldsymbol{\varepsilon}}_e + \partial\tilde{\boldsymbol{g}}_p^T\tilde{\boldsymbol{\Pi}}_p\tilde{\boldsymbol{\varepsilon}}_p + \partial\tilde{\boldsymbol{f}}_i^T\tilde{\boldsymbol{\Pi}}_{x,i}\tilde{\boldsymbol{\varepsilon}}_{x_i} - \mathcal{D}^T\tilde{\boldsymbol{\Pi}}_{x,i}\tilde{\boldsymbol{\varepsilon}}_{x,i} \\ \dot{\tilde{\boldsymbol{\mu}}}_e - \mathcal{D}\tilde{\boldsymbol{\mu}}_e &= -\partial_{\mu_e}\mathcal{F} = -\tilde{\boldsymbol{\Pi}}_e\tilde{\boldsymbol{\varepsilon}}_e + \partial\tilde{\boldsymbol{g}}_v^T\tilde{\boldsymbol{\Pi}}_v\tilde{\boldsymbol{\varepsilon}}_v + \partial\tilde{\boldsymbol{f}}_e^T\tilde{\boldsymbol{\Pi}}_{x,e}\tilde{\boldsymbol{\varepsilon}}_{x_e} - \mathcal{D}^T\tilde{\boldsymbol{\Pi}}_{x,e}\tilde{\boldsymbol{\varepsilon}}_{x,e} \end{aligned} \tag{37}$$

For both units, we note an extrinsic prediction error acting either as a prior or as an observation, a sensory-level observation (either proprioceptive or exteroceptive), and a dynamics prediction error from previous or successive temporal orders. This two-level architecture is highly effective in tasks that require dynamic constraints in both intrinsic and extrinsic domains (e.g., when moving the arm while keeping the hand's palm up) but fails when simultaneous coordination of multiple limbs is needed. In this case, we can extend the model by designing an *intrinsic-extrinsic module* for every degree of freedom of the agent's body. Then, a dynamic attractor at the last (e.g., hand) level results in a prediction error that is backpropagated throughout the whole hierarchy, eventually inferring an appropriate hierarchical configuration of the body.