Cognitive Effort in the Two-Step Task: An Active Inference Drift-Diffusion Model Approach

Álvaro Garrido-Pérez[1][0009−0003−5481−8166], Viktor Lemoine[1], Amrapali Pednekar[1][0009−0005−6194−3955], Yara Khaluf[2][0000−0002−5590−9321], and Pieter Simoens[1][0000−0002−9569−9373]

[1] IDLab, Department of Information Technology, Ghent University - imec, Belgium
`alvaro.garridoperez@ugent.be`
[2] Wageningen University & Research, Wageningen, The Netherlands

Abstract. High-level theories rooted in the Bayesian Brain Hypothesis often frame cognitive effort as the cost of resolving the conflict between habits and optimal policies. In parallel, evidence accumulator models (EAMs) provide a mechanistic account of how effort arises from competition between the subjective values of available options. Although EAMs have been combined with frameworks like Reinforcement Learning to bridge the gap between high-level theories and process-level mechanisms, relatively less attention has been paid to their implications for a unified notion of cognitive effort. Here, we combine Active Inference (AIF) with the Drift-Diffusion Model (DDM) to investigate whether the resulting AIF-DDM can simultaneously capture the effort arising from both habit violation and value discriminability. To our knowledge, this is the first time AIF has been combined with an EAM. We tested the AIF-DDM on a behavioral dataset from the two-step task and compared its predictions to an information-theoretic definition of cognitive effort based on AIF. The model's predictions successfully accounted for second-stage reaction times but failed to capture the dynamics of the first stage. We argue the latter discrepancy likely stems from the experimental design rather than a fundamental flaw in the model's assumptions about cognitive effort. Accordingly, we propose several modifications of the two-step task to better measure and isolate cognitive effort. Finally, we found that integrating the DDM significantly improved parameter recovery, which could help future studies to obtain more reliable parameter estimates.

Keywords: Active inference · Drift-diffusion model · Cognitive effort

1 Introduction

Building upon the Bayesian Brain Hypothesis (BBH), numerous studies in the past decade have tried to formalize cognitive effort using information-theoretic principles [25]. According to the BBH, humans maintain an internal world model encoded in prior beliefs, which they continuously update as they interact with the environment. Within this context, cognitive effort arises from the conflict

between a pre-existing belief about how to act (a habit) and an updated belief about the optimal policy [10].

In a decision-making task, cognitive effort may also arise from the competition between the subjective values of the available choices. When these values are closer together—that is, when value discriminability is low—reaction times (RTs) tend to increase (e.g., [7, 3]). Although RTs are not a direct measure of cognitive effort, they are often used as a proxy (e.g., [24]), based on the assumption that slower responses reflect more information processing.

From an information-theoretic perspective, the effect of value discriminability on RTs can be understood as the additional cognitive effort required to resolve increased choice uncertainty [13, 8, 6]. Yet, information theory provides no explicit account of the underlying deliberation process, which complicates the task of linking its predictions to specific neural signatures. In contrast, evidence accumulator models (EAMs) offer a mechanistic explanation for how value competition shapes decision speed, which may be empirically tested (e.g., [18]).

A major limitation of using EAMs to study cognitive effort is that they are agnostic to how beliefs are formed. This limitation is often addressed by combining EAMs with Reinforcement Learning (RL) [11]. However, in recent years, Active Inference (AIF) has emerged as a powerful alternative to RL, offering a first-principles perspective on perception, learning, and decision-making [5].

Here, we investigate whether integrating a Drift Diffusion Model (DDM), a prominent class of EAM, with AIF can simultaneously capture the influence of both value discriminability and habit violation on cognitive effort. To our knowledge, this is the first attempt to combine AIF with an EAM to model human behaviour. We evaluate the integrated AIF-DDM using the two-step task, a version of the multi-armed bandit in which participants must plan two steps ahead [2]. Furthermore, we compare its predictions with a recently proposed definition of cognitive effort in AIF [10].

Our work builds directly on a recent study that developed an AIF model of the two-step task [4]. The study demonstrated that AIF outperformed a Hybrid Reinforcement Learning (HRL) model (which combines model-free and model-based strategies [16]) in two out of four datasets, while achieving comparable performance in the remaining two. Moreover, the authors provided compelling evidence for directed exploration—a key differentiator between AIF and HRL. Despite these achievements, the study could not determine which specific AIF learning mechanisms participants were using. Given prior evidence that integrating a DDM into an HRL model improved the reliability of model-based estimates [16], we tested whether the combined AIF–DDM could resolve this ambiguity.

2   Methods

2.1   Participants and behavioural task

In this section, we will briefly describe the two-step task and the behavioural dataset that we used to fit the computational models. For more details on the

experimental procedure, we encourage reading the paper that made the dataset publicly available [17].

As the name suggests, each trial of the two-step task consists of two stages (Fig. 1). In the first stage, participants choose between two actions. Each action leads to one of the two second-stage states through a probabilistic transition that is either common ($p = 0.7$) or rare ($p = 0.3$). Importantly, the two first-stage actions have opposite most-likely transitions. These transition probabilities remain constant throughout the task. After transitioning to a second-stage state, participants make a final choice between two actions. Each of these second-stage actions results in a monetary reward or no reward, depending on its current *outcome* probability. In contrast to the fixed transitions, these outcome probabilities fluctuate independently over time, following Gaussian random walks.
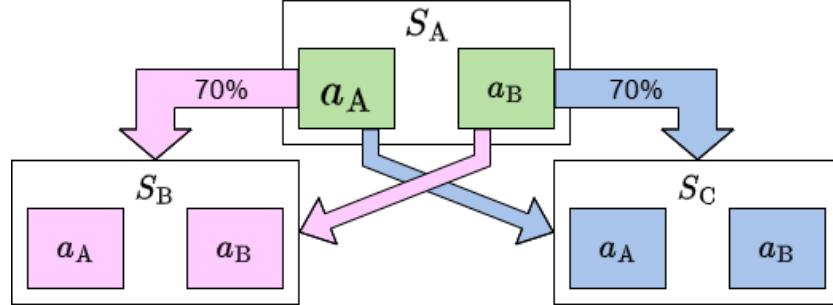

Fig. 1: Abstract representation of the two-step task. At the first stage (top), a choice leads to one of two second-stage states (bottom). Transitions are either common (p=0.7, thick arrows) or rare (p=0.3). The two initial actions have opposing common transitions. A second-stage choice may result in a monetary reward with a probability that fluctuates over time.

We analysed data from the "Magic Carpet" dataset, which was made available by [17] and originally comprised 24 participants. In this experiment, participants had a 2-second deadline to respond at each stage. Every time this deadline was surpassed, the trial was labelled as a *missed* trial and the participant moved on to the next stage or trial. Before conducting any analysis, we identified and removed these trials. We also considered invalid trials those with at least one RT smaller than 100 ms (to exclude anticipatory responses that are too fast to reflect genuine deliberation [9]). Finally, any participant with more than 10% invalid trials was removed from the dataset. One participant was excluded under this criterion, with 44.1% of their trials being either missed or too fast. Therefore, the original sample size was reduced to n=23. In total, 6.53% of the data was excluded from the analysis (including all the data from the removed participant).

2.2   An active inference drift-diffusion model of the two-step task

In this section, we introduce the AIF model developed by [4] and discuss how we integrate it with a DDM. Further explanations of the equations are provided in Appendix A. To establish a comparative benchmark, we implemented an HRL-DDM. Since the latter model is secondary to our main analysis, its mathematical formulation is only detailed in Appendix B.

In what follows, we will use the same notation as in [4]. For a given trial t, the first-stage state and chosen actions are denoted by $s_{1,t}$ and $a_{1,t}$, respectively. Likewise, the second-stage state and chosen action are denoted by $s_{2,t}$ and $a_{2,t}$, respectively. Note that the first-stage state can only be s_A, but the second-stage state may be s_B or s_C, depending on the transition (see Fig. 1). Outcomes $o_t \in \{0, 1\}$ represent the observed reward (or absence of it) after choosing a second-stage action.

According to the AIF model developed by [4], an agent performing the two-step task is equipped with a generative model (i.e., a set of beliefs about how the task environment evolves given its actions). In the two-step task, agents must learn the outcome probabilities of each of the four second-stage actions. We will refer to the set of these four probabilities as θ, and the belief distribution over θ, at a given trial t, $\pi_t(\theta)$. In addition, the agent must learn the four transition probabilities $p(s_{2,t}|s_{1,t}, a_{1,t})$. However, as pointed out by [4], these transitions are accurately learned in a few trials and therefore, action-selection is only sensitive to information regarding outcome probabilities.

Thus, an agent performing the two-step task must optimally balance getting as many rewards as possible (exploitation) and learning the four outcome probabilities (exploration). According to AIF, this balance is achieved by selecting actions that minimize a quantity known as Expected Free Energy (EFE), which in this case is given by the following equation:

$$G_t(a) = \underbrace{-\mathbb{E}_{p(o_t; \pi_t(\theta)|a)}[\ln p(o_t|C)]}_{\substack{\text{Realising preferences} \\ \text{(Exploitation)}}} - \underbrace{\mathbb{E}_{p(o_t; \pi_t(\theta)|a)}[D_{\text{KL}}(\pi_t(\theta)|o_t, a \| \pi_t(\theta))]}_{\substack{\text{Model parameter exploration} \\ \text{(Active Learning)}}} \quad (1)$$

Where $G_t(a)$ is the EFE of a given action at a given trial, $p(o_t|C)$ is the distribution over prior preferred observations (Eq. 5, Appendix A), which depends on a free parameter λ. Note that in the two-step task, the *preferred* observation is to get the reward after a second-stage choice. D_{KL} is the Kullback-Leibler divergence between prior beliefs about second-stage outcome probabilities $\pi_t(\theta)$ and posterior beliefs after selecting an action and observing its outcome. $p(o_t; \pi_t(\theta)|a)$ is a distribution equal to $p(o_t|\theta, a)\pi_t(\theta)$. Note that, unlike traditional EFE formulations, there is no *hidden state exploration* term because in the real experiment participants can always see the state in which they are [15].

Eq. 1 can be used to calculate EFEs for second-stage actions. However, for first-stage actions, the level of optimality depends upon the EFEs of the final states and the transition probabilities. Therefore, for the first-stage actions, the EFE equation is given by:

$$G(a_j) = p(s_B|s_A, a_j) \sum_{a_2 \in \mathcal{A}_B} G(a_2) \; + \; p(s_C|s_A, a_j) \sum_{a_2 \in \mathcal{A}_C} G(a_2) \qquad (2)$$

Where $\mathcal{A}_B$ and $\mathcal{A}_C$ are the sets of available actions in the second-stage states s_B and s_C respectively, and $G(a_2)$ are the second-stage EFEs given by Eq. 1.

Previous studies have reported a tendency for participants to repeat initial-stage actions, regardless of the outcome history [4]. Within AIF, this behaviour is formalised through a habit term E. In the model of [4], E is parameterised by κ, which modulates the extent of an agent's reliance on habits (see Eq. 6, Appendix A).

At the first stage, an agent will select an action that minimizes the EFE corrected by the habitual bias $G_t^{\text{net}}(s_1, a_1) = G_t(s_1, a_1) + E(a_1)$. At the second stage, no habitual bias is assumed, and therefore $G_t^{\text{net}}(s_2, a_2) = G_t(s_2, a_2)$.

For the original AIF model [4], the choice probabilities of an agent at trial t, stage $p = \{1, 2\}$ and state s, will be given by the softmax distribution of the corresponding net EFEs parametrized by an inverse temperature parameter, γ_p (one for each stage). However, for the AIF-DDM the choice will be determined by a drift-diffusion process, with a non-decision time t_{nd}, boundary separation a_{bs} and a drift rate $v_{p,s,t}$, that depends on the difference in G^{net} of the two available actions such that:

$$v_{p,s,t} = v_p^{\text{mod}}[G_t^{\text{net}}(s, a) - G_t^{\text{net}}(s, a')], \;\; a, a' \in \mathcal{A}_s \qquad (3)$$

Where $\mathcal{A}_s$ is the set of available actions at state s, and v_p^{mod} is a free parameter that regulates the sensitivity to G^{net} differences (which can also be interpreted as the agent's information-processing speed). Following the original AIF model [4], we fit a separate v_p^{mod} parameter for each stage, an approach analogous to the use of stage-specific inverse temperatures.

In the classical DDM, a starting-point bias parameter z is often included. However, in the two-step task, the symbols representing the different actions were randomly displayed either on the left or the right of the screen [17]. Therefore, we set $z = 0.5$ (effectively cancelling the bias in the drift-diffusion process).

After completing every trial, the agent updates its beliefs about the second-stage outcome probabilities. In [4], the updating rules for the prior distribution over outcome probabilities (Eqs. 8 and 9 in Appendix A), depend on four free parameters: the learning rate l, the prior volatility, v_{PS}, which modulates the influence of surprise on the agent's beliefs, volatility of sampled actions v_{SD}, which modulates how beliefs over chosen actions *decay*, or are forgotten, and the volatility of unsampled actions v_{UD}, similar to v_{SD} but for unchosen actions.

The study by [3] tested four AIF variants distinguished by their learning rules: a No Unsampled-Decay model (AIF$_{\text{NUD}}$, $v_{\text{UD}} = 0$); a No Sampled-Decay model (AIF$_{\text{NSD}}$, $v_{\text{SD}} = 0$), a No Predictive Surprise model (AIF$_{\text{NPS}}$, $v_{\text{PS}} = 0$) and a model including all the learning mechanisms (AIF$_{\text{FULL}}$). Although solid evidence was found favouring AIF over HRL, the best-fitting AIF variant could not be determined [4]. Consequently, we fitted all four models in our analysis.

2.3   Cognitive effort and active inference

A recent formalisation of cognitive effort (ξ) based on AIF defines it as the KL divergence between context-sensitive beliefs about how to act $P_G(\pi)$ and context-insensitive prior beliefs, or habits $P_E(\pi)$ [10]:

$$\underbrace{\xi \triangleq D_{\mathrm{KL}}[P_G(\pi)||P_E(\pi)]}_{\text{Effort}} = \underbrace{\mathbb{E}_{P_G}[\ln P_G(\pi)]}_{\text{Context sensitive}} - \underbrace{\mathbb{E}_{P_G}[\ln P_E(\pi)]}_{\text{Context insensitive}} \qquad (4)$$

$$P_E(\pi) = \mathrm{Cat}(\sigma(-\mathbf{E})) \qquad P_G(\pi) = \mathrm{Cat}(\sigma(-\mathbf{G} - \mathbf{E}))$$

Where $\mathbf{G}$ and $\mathbf{E}$ are vectors comprising the context-sensitive EFEs and the context-insensitive priors (or habits) of the available actions, respectively.

From Eq. 4, two key predictions can be derived. First, high effort will be experienced when there is an incongruence between $\mathbf{G}$ and $\mathbf{E}$. Second, when the elements of $\mathbf{G}$ are of similar magnitude to one another, cognitive effort is minimal regardless of $\mathbf{E}$.

For the first stage of the two-step task, the AIF-DDM aligns with the first prediction of Eq. 4. For a choice between actions a_A and a_B, with $\mathbf{G} = [G_A, G_B]$, and $\mathbf{E} = [E_A, E_B]$, RTs should increase as the magnitude of the drift rate decreases, which is proportional to $|\Delta G^{\mathrm{net}}| = |G_A + E_A - G_B - E_B|$. Therefore, higher RTs may occur when, for example $|G_A| \gg |G_B|$, $|E_A| \ll |E_B|$ and $|G_A| \approx |E_B|$. In other words, when there is an incongruence between $\mathbf{G}$ and $\mathbf{E}$.

However, our model contradicts the second prediction of Eq. 4, because for a constant $\mathbf{E}$, RTs should increase as $|G_A|$ gets closer to $|G_B|$. Interestingly, for the second stage, Eq. 4 predicts minimal effort if we assume no habits (i.e. $E_A = E_B = 0$), in contrast to AIF-DDM, where effort can still be high if the context-sensitive EFEs are closely matched.

2.4   Model fitting and comparison procedures

We followed the same Maximum Likelihood Estimate (MLE) procedure as in [4], to fit the models' free parameters to the behavioural data [3]. For each participant and model, we found the parameter set with the highest likelihood using Scipy's 'L-BFGS-B' algorithm [19]. This step was repeated 35 times for each participant, with different (uniformly) randomized initializations for all parameters. After completing all the runs, we selected the parameter set with the maximum likelihood and used its value for model comparison.

For the pure AIF and HRL models, we computed the likelihoods using the choice probability distributions of each model. For AIF-DDM and HRL-DDM however, the likelihoods were given by the Wiener's First-Passage Time Distribution (WFPT) provided by the HDDM Python package [22]. Further details can be found in Appendix C.

---

[3] **Code** available at https://github.com/decide-ugent/aif-ddm

For model comparison, we relied on the Statistical Parametric Mapping (SPM) [21] and Variational Bayesian Analysis (VBA) [1] toolboxes, using MAT-LAB R2022b. We performed a group-level random-effect Bayesian model selection (BMS) procedure, which requires the log-model evidence (LME) of each model for each participant. We used two LME approximations: the Bayesian Information Criterion (BIC) and the Akaike's Information Criterion (AIC) scores [12] (see Appendix C). Using the BMS procedure, we estimated four model comparison metrics: the Expected Posterior Probability (EPP); the Estimated Model Frequency (EMF), or the proportion of participants best fit by the model; the Exceedance Probability (EP), which quantifies the likelihood that the model is more frequent than competing models across participants; and the Protected Exceedance Probability (PEP), similar to EP but accounting for the possibility that apparent differences in model frequencies arise due to chance [14].

The SPM toolbox allows comparing the performance of model *families* (i.e., models that share a common feature). As in [4], we use this feature to compare the overall AIF performance (considering its four variants) to HRL. Since PEPs are unavailable for family-level inference, we only calculate EPs for these analyses.

2.5   Model and parameter recovery procedures

To perform the parameter recovery analysis, we simulated a dataset for each model using parameter values sampled from uniform distributions with ranges equal to those used to fit the real dataset. Next, we fitted the synthetic dataset and obtained new (recovered) parameters. Finally, we checked if the original and recovered parameters were correlated by calculating Pearson's correlations. This process was repeated 23 times (equal to the number of participants in the real experiment) for each model.

For the model recovery analysis, each model was first used to simulate a full experimental dataset with the same parameter sets as in the parameter recovery analysis. We then fitted all models to each simulated dataset and calculated the proportion of times each model was identified as best-fitting. Results were summarized in a confusion matrix (Fig. 2).

3   Results

3.1   Model recovery analysis

Model recovery analysis indicates that AIF-DDM$_{\text{FULL}}$ is frequently misclassified as either AIF-DDM$_{\text{NSD}}$ or AIF-DDM$_{\text{NSD}}$ (Fig. 2). The poor recovery likely stems from the full model encompassing all update rules present in the other two models [4]. The rest of the models had a better but modest recovery (except HRL-DDM, which had a perfect recovery). Thus, combining AIF with a DDM could not help substantially to differentiate between different learning variants as we hypothesised. Overall, both AIC and BIC scores show similar results. Although the AIC score seems slightly more reliable since it could achieve a better recovery for both AIF-DDM$_{\text{FULL}}$ and AIF-DDM$_{\text{NUD}}$.
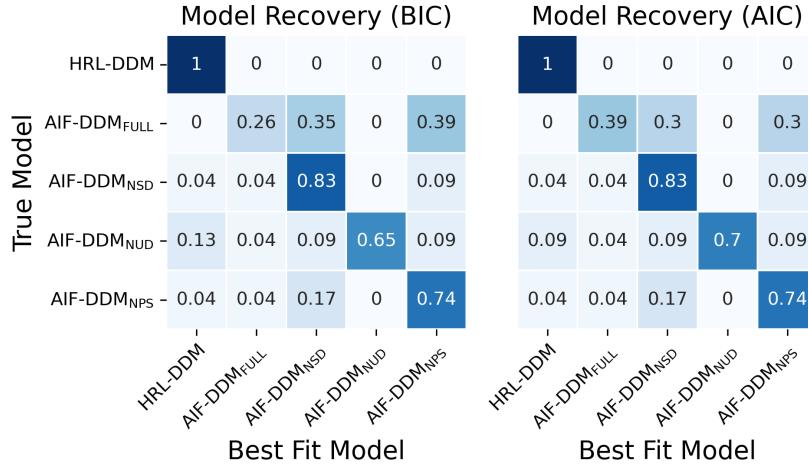
Fig. 2: Model recovery results. Each cell contains the fraction of simulated experiments from a given true model, for which a corresponding model was classified as the best-fitting, according to the BIC score (left) or the AIC score (right).

3.2   Model comparison results

The BMS analysis revealed different results depending on the score used. According to the BIC score, AIF-DDM$_{NSD}$ and AIF-DDM$_{NPS}$ dominated moderately across participants with Expected Posterior Probabilities, EPP $= [0.39, 0.31]$, Protected Exceedance Probabilities, PEP $= [0.63, 0.29]$ and Estimated Model Frequencies, EMF $= [0.44, 0.35]$, respectively (see Fig. 5, first row). However, according to the AIC scores AIF-DDM$_{FULL}$ was the best-fitting model with EPP $= 0.57$, PEP $= 0.97$ and EMF $= 0.7$ (see Fig. 5, second row). The disagreement between AIC and BIC scores was also reported for the model comparison between pure HRL and AIF models [4], and likely stems from the fact that BIC penalizes complexity more than AIC.

Even though the two BMS procedures produced conflicting results, we selected AIF-DDM$_{FULL}$ as the best-fitting model, and focus exclusively on its results in the subsequent sections. This decision was based on the model recovery analysis, which showed that the full model is frequently misclassified as either AIF-DDM$_{NSD}$ or AIF-DDM$_{NPS}$. Nonetheless, we acknowledge that the BIC-based comparison favoured AIF-DDM$_{NSD}$ and AIF-DDM$_{NPS}$, so we replicated our analysis for these models and show their results in Appendix D. We excluded AIF-DDM$_{NUD}$ from further analysis as it performed poorly in both BMS procedures. Finally, the model family selection analysis revealed similar results to those reported in the study that compared pure AIF and HRL [4]. We found that the AIF-DDM family outperformed the HRL-DDM model according to both BIC (EP $= 1$, EMF $= 0.9$, EPP $= 0.88$) and AIC (EP $= 1$, EMF $= 0.92$, EPP $= 0.9$) scores.

3.3   Parameter recovery analysis

Integrating a DDM with AIF substantially improved the recovery of the parameters shared between AIF-DDM and pure AIF models (see Appendix E, Table 1). For the pure AIF models, parameter recovery was far from perfect. For example, for $\mathrm{AIF_{FULL}}$, only v_{DU} had a Pearson's correlation greater than .6 between its ground-truth and recovered values. In contrast, for the AIF-DDM models, v_{DU}, v_{DS}, v_{PS}, a_{bs} and t_{nd} showed excellent recovery ($r > .90$, $p < .001$), l and κ were well recovered ($r > .84$, $p < .001$) and p_r and λ had a moderate recovery ($r > .67$, $p < .001$). The drift rate parameters v_1^{mod} and v_2^{mod} had moderate-to-poor recovery, depending on the specific model. Nonetheless, their recovery was still superior to that of the inverse temperature parameters (γ_1 and γ_2) from the pure AIF model.

3.4   Expected free energy discriminability affects second stage, but not first stage reaction times

To evaluate the models' goodness-of-fit and predictive accuracy, we compared model-simulated behaviour on the two-step task to the observed behaviour of participants. This was done using a decile-binned analysis based on the absolute difference in the net EFE between the two actions, $|\Delta G^{\mathrm{net}}|$. For each AIF-DDM model (except $\mathrm{AIF\text{-}DDM_{NUD}}$), we first computed the trial-by-trial $|\Delta G^{\mathrm{net}}|$ for each participant for both stages, based on their best-fit model parameters, choice history, and the experienced sequence of rewards and transitions. From the resulting distribution of $|\Delta G^{\mathrm{net}}|$ values, for each participant and stage, we identified the decile boundaries (10th-90th percentiles). We then binned each trial's observed RT and choice into one of ten decile bins according to its $|\Delta G^{\mathrm{net}}|$ value. For example, a trial with a $|\Delta G^{\mathrm{net}}|$ value below the 10th percentile was placed in the first bin (0th-10th). Within each bin, we calculated two metrics for each participant: (1) the mean RT, and (2) the probability of choosing the action with the lower net EFE, $P(\text{choose } G_{\mathrm{min}}^{\mathrm{net}})$. These participant-level metrics were then averaged across participants for each decile bin. Finally, for each trial stage, we generated 100 simulations to estimate the model's predicted RT and choice probability. These simulated metrics were then binned and averaged in the same way as the observed data.

The simulation analysis results were qualitatively similar across all models, therefore we only display the predictions of the $\mathrm{AIF\text{-}DDM_{FULL}}$ model as an example in Fig. 3. Equivalent plots for the remaining models are presented in the Appendix F. In both stages, participants were more likely to select the action with the minimum net EFE as the value of $|\Delta G^{\mathrm{net}}|$ increased. This trend was well-captured by all the AIF-DDM models (see Appendix F).

The AIF-DDM models also predict that the mean RT should decrease as a function of EFE discriminability (i.e., as $|\Delta G^{\mathrm{net}}|$ increases) for both stages. This effect can be observed for the second stage; however, it is almost negligible for the first one (Fig. 3, bottom left).
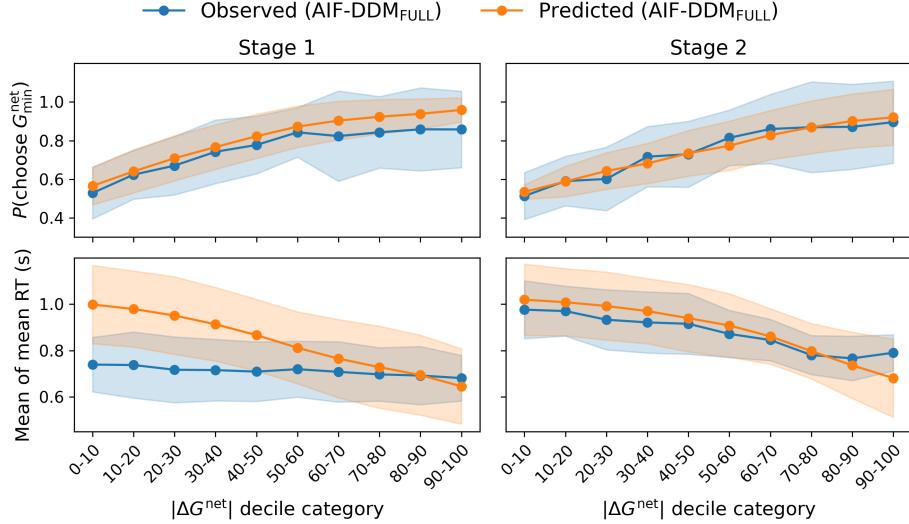
Fig. 3: Comparison of observed (blue) and model-predicted (orange) behaviour. Lines show the mean choice probability (top) and mean reaction time (bottom) across participants, binned by net Expected Free Energy ($|\Delta G^{\text{net}}|$) deciles. Shaded areas are ± 1 standard error of the mean.

Multiple explanations may account for the mismatch in first-stage RTs. The first concerns the 2-second deadline. As described in Section 2.1, all missed trials were excluded from the analysis, effectively truncating the observed RT distribution. In contrast, the AIF-DDM models do not impose this time restriction, resulting in higher predicted RT means (see Appendix G for further details). However, this effect alone is unlikely to fully explain the misfit, since the second stage had the same deadline, yet a far less pronounced mismatch.

A second possible explanation stems from the well-documented tendency for participants to repeat their previous first-stage choice [4]. In both AIF(-DDM) and HRL(-DDM), a habit term is included only in the first-stage equations. It is therefore plausible that, for most first-stage choices, the habit strength of the previously selected option is sufficient to produce fast responses—even in trials falling into the lowest $|\Delta G^{\text{net}}|$ deciles. In contrast, with the lack of a habit term in the second-stage equations, $|\Delta G^{\text{net}}|$ values may be smaller or more variable. This effect was particularly notable for the participant with the largest fitted κ value (where κ regulates the tendency to repeat the previous first-stage choice). However, it was not observed for the rest of the participants (see Fig. 4).

A third potential explanation is that a substantial portion of the cognitive effort in the initial stage may arise from mental simulation or partial exploration of the decision tree, a process not accounted for by the assumption that RTs depend solely on $|\Delta G^{\text{net}}|$. Nevertheless, the limited complexity of the task (involving only four possible branches) constrains the extent of such planning.
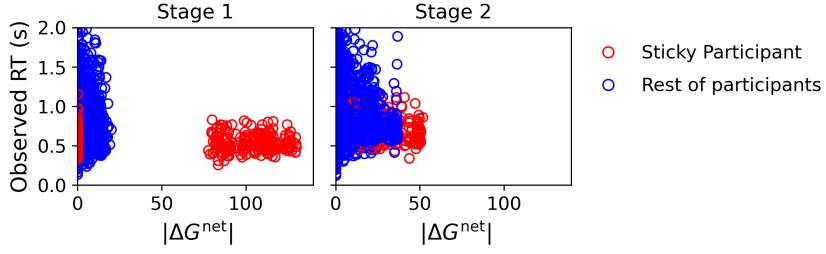
Fig. 4: Observed reaction time (RT) vs. $|\Delta G^{\mathrm{net}}|$ for the AIF-DDM$_{\mathrm{FULL}}$ model, separated by task stage. Each circle represents a single trial. Trials from the 'sticky' participant, identified by the highest fitted κ value, are shown in red. Trials from all other participants are shown in blue. For the sticky participant, notice the cluster of first-stage trials with high $|\Delta G^{\mathrm{net}}|$ values compared to the second-stage trials that are more concentrated at low $|\Delta G^{\mathrm{net}}|$ values.

3.5   The lack of net EFE discriminability effect on first-stage reaction times might be explained by the experimental design

A final potential explanation for the lack of $|\Delta G^{\mathrm{net}}|$ effect on first-stage RTs is the experimental design. In the trial sequence of the two-step task, the elapsed time from the moment participants make the second-stage choice of the previous trial until they are presented with the first-stage options of the next trial is variable and long (3.2-3.8s). This time interval could allow participants to engage in "pre-thinking" about their upcoming first-stage choice before the options are presented. As a result, the recorded first-stage RT may only capture the final execution of an already made decision, rather than the full deliberation process. In contrast, the interval between the first- and second-stage choices is fixed and shorter (2s), leaving less room for pre-thinking and making the second-stage RT a more reliable measure of cognitive effort.

The difference in inter-choice intervals could also explain why mean first-stage RTs are shorter than mean second-stage RTs (see Fig. 3), which may seem counterintuitive since, according to AIF, participants in the first stage should engage in planning, a more costly cognitive process than the one-shot second-stage decision.

4   Discussion

In this study, we introduced a novel AIF-DDM and evaluated its predictions in the two-step task. We discussed the difference between the model predictions and a recently developed information-theoretic formalism of cognitive effort based on AIF (Eq. 4, [10]).

Our model predicts that both the need to overcome a pre-existing habit and the difficulty of discriminating between options with similar subjective values (EFEs) increase cognitive effort during the first stage of the two-step task. We

argue that for the first stage, both effects can be regarded as a single competition process between the 'habit-corrected' or net EFEs of the available options. For the second stage, habits are assumed to play no role, and therefore, the model predicts that cognitive effort only depends on the discriminability between pure context-sensitive EFEs.

We found empirical evidence that EFE discriminability affects second-stage RTs of the two-step task as predicted by the AIF-DDM model – an effect that is not captured by the information-theoretic formalism. However, we found that the discriminability between the habit-corrected EFEs had a negligible effect on the first-stage RTs, contrary to the AIF-DDM predictions. We argue that the latter observation could be a result of the long inter-trial intervals that may allow participants to "pre-think" their choice before being presented with the options. Thus, first-stage RTs may only reflect motor execution, not the preceding cognitive effort.

Lastly, we found that integrating a DDM with AIF did not help to distinguish between different learning mechanisms proposed by [4]. However, the DDM integration improved parameter recovery compared to pure AIF. The latter result was expected, since the same effect has been reported for HRL in the two-step task [16]. Nonetheless, the result might be relevant for computational psychiatry studies interested in finding the link between certain AIF parameters and specific pathologies.

Our study has multiple limitations. Firstly, we focused on comparing the predictions of AIF-DDM with a single information-theoretic definition of cognitive effort. Our intention, however, is not to disprove this particular formalism, but to illustrate that a more complete theory of cognitive effort should include both the effect of subjective value competition and the violation of prior expectations. Secondly, we focused on studying how cognitive effort may result from the competition between the EFEs and prior beliefs of the different available options. We did not, however, address the costs associated with learning or with *generating* EFEs, which are likely to differ between the two stages. Finally, cognitive effort is a complex phenomenon that may depend upon multiple factors, such as the objective cognitive demand of a task or the participant's information-processing speed. Some of these factors may be captured by different AIF-DDM parameters (e.g., a_{bs} and κ for impulsiveness and motivation, respectively); however, this analysis was beyond the scope of this paper.

A future study could evaluate whether the pre-thinking effect is responsible for the AIF-DDM misfit for the first-stage RTs by reducing the inter-trial intervals. This could be achieved by, for example, reducing the amount of time that the rewards are displayed. Moreover, the problem of ruling out missed trials could be mitigated by increasing the available time to make a choice, using a likelihood function that can take into account the probability of missing a deadline, or introducing collapsing thresholds in the DDM [20]. Implementing these changes could move us closer to a better understanding of cognitive effort in the two-step task.

A   Further details on Active Inference in the two-step task

In [4], the distribution over prior preferred observations is given by the following equation:

$$P(o_t|C) = \frac{1}{Z(\lambda)}e^{o_t\lambda}e^{-(1-o_t)\lambda} \tag{5}$$

Where λ is a free parameter that regulates how much an agent prefers to realize prior preferences (in this case, getting a reward) over gaining new information. Thus, λ regulates the exploration-exploitation trade-off.

As mentioned in Section 2.2, agents are assumed to tend to repeat first-stage actions independent of the observed outcomes. This behaviour is modelled in AIF by introducing a *habit* term. In [4], this quantity is described by:

$$E(a_j) = \frac{1}{Z(\kappa)}e^{\delta_{a_{t-1},a_j}\kappa_e - (1-\delta_{a_{t-1},a_j})\kappa} \tag{6}$$

Where κ is a precision parameter regulating the agent's reliance on habits.

After every trial, the AIF agent updates its beliefs about the second-stage outcome probabilities. Since these probabilities are mutually independent, the overall distribution equals the product of four beta distributions, each describing the believed outcome probability of its respective action and parameterized by its corresponding α and β.

$$\pi_t(\theta) = \prod_{s=1}^{2}\prod_{a=1}^{2}\mathcal{B}e(\alpha_{s,a,t}, \beta_{s,a,t}) \tag{7}$$

According to the learning rule implemented by [4], after observing the outcome of a chosen action, its corresponding α and β parameters are updated following these rules:

$$\alpha_{s,a,t} = (1-\chi_t)\alpha_{s,a,t-1} + \delta_{a_t,a}o_t l + (1-\nu_{SD})\alpha_{s,a,t-1} + \nu_{SD}\alpha_0 \tag{8}$$

$$\beta_{s,a,t} = (1-\chi_t)\beta_{s,a,t-1} + \delta_{a_t,a}(1-o_t)l + (1-\nu_{SD})\beta_{s,a,t-1} + \nu_{SD}\beta_0 \tag{9}$$

$$\chi_t = \frac{m\,PS}{1+m\,PS} \quad \text{(10a)} \qquad\qquad m = \frac{\nu_{\mathrm{PS}}}{1-\nu_{\mathrm{PS}}} \quad \text{(10b)}$$

Where $\chi_t \in [0,1]$ is the surprise-modulated adaptation rate for trial t which depends on the predictive surprise $PS = -\ln p(o_t; \pi_t(\theta))$, and the prior volatility parameter $\nu_{\mathrm{PS}} \in [0,1]$ which modulates the influence of surprise on the agent's beliefs. In Eqs. 8 and 9, l is the learning rate parameter, ν_{SD} is the volatility parameter of chosen actions and ν_{SD} modulates how over time α and β decay towards their prior values α_0 and β_0, respectively.

Evidence from multi-armed bandit tasks suggests a bias in prior outcome probabilities. To model this effect, a free parameter, p_r, is fitted to each participant, which determines the initial (or uninformed) α and β prior values, such that $\alpha_0 = 2(1 - p_r)$ and $\beta_0 = 2p_r$ [4].

The decay effect of the chosen actions captured by ν_{SD}, may also affect unchosen actions, and therefore, for these, the updating rules reduce to:

$$\alpha_{s,a,t} = (1 - \nu_{\mathrm{UD}})\alpha_{s,a,t-1} + \nu_{\mathrm{UD}}\alpha_0 \tag{11}$$

$$\beta_{s,a,t} = (1 - \nu_{UD})\beta_{s,a,t-1} + \nu_{UD}\beta_0 \tag{12}$$

Where ν_{UD} is the volatility of unchosen actions.

The surprise update mechanism for chosen actions and the decay mechanisms for chosen and unchosen actions may not all be simultaneously operating [4]. Therefore, four variants of the AIF models can be implemented, each with a different combination of update rules as pointed out in Section 2.2.

In addition to the outcome probabilities, agents must also learn the transition probabilities. As pointed by [4], before starting the experiment, participants completed an instruction phase as well as 50 practice trials and therefore, they were aware that the transition probabilities of initial-stage actions were mirrored and could either be $p(s_B|s_A, a_A) = p(s_C|s_A, a_B) = 0.7$ or $p(s_B|s_A, a_A) = p(s_C|s_A, a_B) = 0.3$ (with $p(s_B|s_A, a_A) = 1 - p(s_C|s_A, a_A)$ and $p(s_B|s_A, a_B) = 1 - p(s_C|s_A, a_B)$, respectively). In [4], this transition learning is modelled by having agents count transitions, and on each trial, choosing the most likely transition structure based on the observed frequencies.

B   Hybrid reinforcement learning drift-diffusion model

According to the HRL-DDM model, an agent performing the two-step task in state s, and trial t, will select an action a that maximizes the expected discounted future reward (or state-action value), denoted by $Q_t(s, a)$. Traditionally, state-action values can be computed using either a model-free Q_t^{MF} or a model-based strategy Q_t^{MB}.

In the two-step task, after completing each trial, model-free state-action values are updated following the SARSA(λ) algorithm. For the chosen second-stage action a_2 and the final state (either s_B or s_c, depending on the transition), its corresponding state-action value is updated as follows:

$$Q_{t+1}^{\mathrm{MF}}(s_2, a_2) = Q_t^{\mathrm{MF}}(s_2, a_2) + \alpha_2\delta_{2,t} \tag{13}$$

Where $\alpha_2 \in [0, 1]$ is the second-stage learning rate parameter and $\delta_{2,t}$, is the second-stage prediction error, defined as the difference between the observed outcome and its state-action value, such that $\delta_{2,t} = o_t - Q_t^{\mathrm{MF}}(s_2, a_2)$

Since there is no outcome after choosing a first-stage action, and its optimality ultimately depends on the outcome observed in the second stage, the updating rule for the model-free state action value of the first-stage action will

depend on the second-stage prediction error $\delta_{2,t}$ and will be given by the following equation:

$$Q_{t+1}^{\mathrm{MF}}(s_{\mathrm{A}}, a_1) = Q_t^{\mathrm{MF}}(s_{\mathrm{A}}, a_1) + \alpha_1[\delta_{2,t} + \lambda\delta_{1,t}] \tag{14}$$

Where $\alpha_1 \in [0,1]$ is the first-stage learning rate parameter, $\delta_{1,t}$, is the first-stage prediction error, defined as $\delta_{1,t} = Q_t^{\mathrm{MF}}(s_2, a_2) - Q_t^{\mathrm{MF}}(s_{\mathrm{A}}, a_1)$, and $\lambda \in [0,1]$ is the eligibility parameter, which determines how much the second-stage prediction error influences the first-stage state-action value.

For the model-based strategy, agents are assumed to have a generative model of the task, which in this case means that they know (or learn) the state transition probabilities and use these to calculate state-action values. For the first-stage actions, the model-based values depend on the second-stage model-free values, such that:

$$Q_t^{\mathrm{MB}}(s_{\mathrm{A}}, a_1) = \sum_{s_2 \in \{s_{\mathrm{B}}, s_{\mathrm{C}}\}} P(s_2|s_{\mathrm{A}}, a_1) \max_{a_2 \in \mathcal{A}_{s_2}} Q_t^{\mathrm{MF}}(s_2, a_2) \tag{15}$$

In the first stage, the agent will follow a combination of model-free and model-based strategies, such that the total net expected reward of a given action a_j, at the first-stage state, s_A is given by a linear combination of the trial's model-free Q_t^{MF} and model-based Q_t^{MB} expected rewards:

$$Q_t^{\mathrm{net}}(s_{\mathrm{A}}, a_j) = wQ_t^{\mathrm{MB}}(s_{\mathrm{A}}, a_j) + (1-w)Q_t^{MF}(s_{\mathrm{A}}, a_j) + \rho\,\mathrm{rep}_t(a_j) \tag{16}$$

Where $w \in [0,1]$ is a weight parameter that regulates the relative influence of model-based and model-free strategies in the decision-making. The last term in the equation is added to model the tendency observed in participants to repeat the first-stage choice of the previous trial. The level of response stickiness is regulated by the parameter ρ, which multiplies the function $\mathrm{rep}(a)$ – equal to 1 if the first-stage action is the same as the one chosen in the previous trial and 0 otherwise.

At the second stage, the state-action values are computed using only a model-free strategy, and it is assumed that there is no tendency to repeat the previous action. Therefore, for the second stage, the net state-action value reduces to $Q_t^{\mathrm{net}}(s_2, a_2) = Q_t^{\mathrm{MF}}(s_2, a_2)$.

Once state-action values are computed, the HRL-DDM agent, will pick an action a at trial t, state s and stage $p = \{1, 2\}$ through a drift-diffusion process with a drift rate given by the following equation:

$$v_{p,s,t} = v_p^{\mathrm{mod}}[Q_t^{\mathrm{net}}(s,a) - Q_t^{\mathrm{net}}(s,a')], \;\; a, a' \in \mathcal{A}_s \tag{17}$$

Where $\mathcal{A}_s$ is the set of available actions at state s.

C   Further details on the model fitting and comparison procedures

For the two-step task, the MLE procedure consists of finding the participant's set of parameter values of the model m, $\hat{\theta}_m^{\mathrm{MLE}}$, that maximizes the likelihood of the first-stage and second-stage behavioural data ($d_{1:T}^1$ and $d_{1:T}^2$, respectively), given the parameters set, $p(d_{1:T}^1, d_{1:T}^2 | \theta_m, m)$, or equivalently, log-likelihood, $LL = \log p(d_{1:T}^1, d_{1:T}^2 | \theta_m, m)$. This quantity can be expressed as the sum of the individual trial LLs [23], such that:

$$LL = \sum_{t=1}^{T} \log p(d_t^1, d_t^2 \mid d_{1:t-1}^1, d_{1:t-1}^2, \theta_m, m) \tag{18}$$

To use a standard minimization procedure, instead of maximizing the LL, we minimize the negative of it (NLL). Moreover, we can express the LL of each trial's data as the product of the first-stage and second-stage data, giving the following expression for the NLL:

$$NLL = -\sum_{t=1}^{T} \log p(d_t^1 \mid d_{1:t-1}^1, d_{1:t-1}^2, \theta_m, m) + \log p(d_t^2 \mid d_{1:t}^1, d_{1:t-1}^2, \theta_m, m) \tag{19}$$

Where $p(d_t^1 \mid d_{1:t-1}^1, d_{1:t-1}^2, \theta_m, m)$ and $p(d_t^2 \mid d_{1:t}^1, d_{1:t-1}^2, \theta_m, m)$ are the probabilities of each choice (or choice and RT) of the first and second stages, respectively, given the parameters of the model and the information available up to that moment.

For AIF and HRL models, d_t^p, corresponds to the choice made at stage p and trial t, so Eq. 19 can be evaluated using the corresponding choice probability distributions of each model. For AIF-DDM and HRL-DDM, however, d_t^p, corresponds to the choice and RT tuple of stage p and trial t. Hence, to compute the likelihood of each trial and phase, we use the WFPT distribution as implemented by the HDDM Python package. The WFPT distribution takes as input the parameters a_{bs}, t_{nd}, as well as the drift rates, which for HRL-DDM and AIF-DDM are calculated using Eqs. 17 and 3, respectively.

To perform the BMS procedure for model comparison, we require an approximation for the LME of each model for each participant. Instead of calculating this quantity, we use two approximations; one based on the BIC score ($LME \approx -BIC/2$) and another based on the AIC score ($LME \approx -AIC/2$) [12], where the BIC and AIC scores are computed using the following equations:

$$\mathrm{BIC} := k \ln(n) - 2\hat{LL} \tag{20}$$

$$\mathrm{AIC} := 2k - 2\hat{LL} \tag{21}$$

Where k is the number of free parameters in the model, n is the number of trials, and $\hat{LL}$ is the maximum LL value, for a given participant and model.

D   Model comparison plots
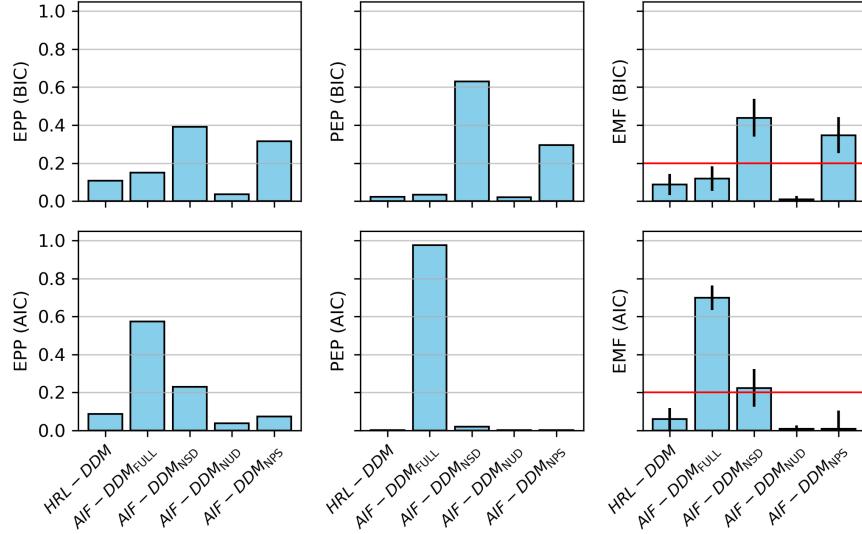
Fig. 5: Bayesian model selection results comparing the five candidate models. The analysis was performed separately using log-model evidence approximations based on BIC (top row) and AIC (bottom row) scores. Each column represents a different metric for model comparison: (left) Expected Posterior Probabilities (EPP), (middle) Protected Exceedance Probabilities (PEP), and (right) Estimated Model Frequencies (EMF). In the EMF plots, the red horizontal line indicates the chance level (0.2), and error bars represent the standard deviation. The AIC-based analysis consistently favours the AIF-DDM$_{\text{FULL}}$ model, while the BIC-based results are more distributed across AIF-DDM$_{\text{NSD}}$, and AIF-DDM$_{\text{NPS}}$.

E   Parameter recovery results

Table 1: Parameter recovery results for the four AIF variants. The Pearson's correlation and p-values between original and recovered parameters for the pure AIF models are denoted by r and p, respectively. For the AIF-DDM models, the Pearson's correlations and p-values are denoted by r_{ddm} and p_{ddm}, respectively.

(a) AIF(-DDM)$_{\mathrm{FULL}}$

Parameter	r	p	r_{DDM}	p_{DDM}
l	.41	.054	.89	$< .001$
v_{DU}	.91	$< .001$	$> .99$	$< .001$
v_{DS}	.42	.043	.99	$< .001$
v_{PS}	.52	.012	.98	$< .001$
λ	.51	.014	.71	$< .001$
κ	.53	.009	.97	$< .001$
p_r	.28	.190	.93	$< .001$
γ_1	.17	.428		
γ_2	.38	.073		
v_1^{mod}			.46	.027
v_2^{mod}			.73	$< .001$
a_{bs}			.98	$< .001$
t_{nd}			$>.99$	$< .001$

(b) AIF(-DDM)$_{\mathrm{NSD}}$

Parameter	r	p	r_{DDM}	p_{DDM}
l	.62	.002	.96	$< .001$
v_{DU}	.90	$< .001$	.97	$< .001$
v_{PS}	.76	$< .001$	.98	$< .001$
λ	.49	.017	.68	$< .001$
κ	.68	$< .001$	.85	$< .001$
p_r	.83	$< .001$	.97	$< .001$
γ_1	.10	.644		
γ_2	.31	.143		
v_1^{mod}			.68	$< .001$
v_2^{mod}			.59	.003
a_{bs}			.99	$< .001$
t_{nd}			$>.99$	$< .001$

(c) AIF(-DDM)$_{\mathrm{NUD}}$

Parameter	r	p	r_{DDM}	p_{DDM}
l	.47	.025	.84	$< .001$
v_{DS}	.69	$< .001$	.93	$< .001$
v_{PS}	.57	.005	.94	$< .001$
λ	.66	$< .001$	.79	$< .001$
κ	.82	$< .001$	.93	$< .001$
p_r	.32	.131	.69	$< .001$
γ_1	.06	.798		
γ_2	.22	.315		
v_1^{mod}			.55	.007
v_2^{mod}			.60	.003
a_{bs}			.95	$< .001$
t_{nd}			$>.99$	$< .001$

(d) AIF(-DDM)$_{\mathrm{NSL}}$

Parameter	r	p	r_{DDM}	p_{DDM}
l	.69	$< .001$	.93	$< .001$
v_{DU}	.94	$< .001$	.91	$< .001$
v_{DS}	.81	$< .001$	.97	$< .001$
λ	.56	.005	.67	$< .001$
κ	.68	$< .001$	.87	$< .001$
p_r	.94	$< .001$	.93	$< .001$
γ_1	.18	.413		
γ_2	.31	.146		
v_1^{mod}			.25	.246
v_2^{mod}			.58	.004
a_{bs}			$> .99$	$< .001$
t_{nd}			$> .99$	$< .001$

F   AIF-DDM$_{\mathrm{NSD}}$ and AIF-DDM$_{\mathrm{NPS}}$ simulation results
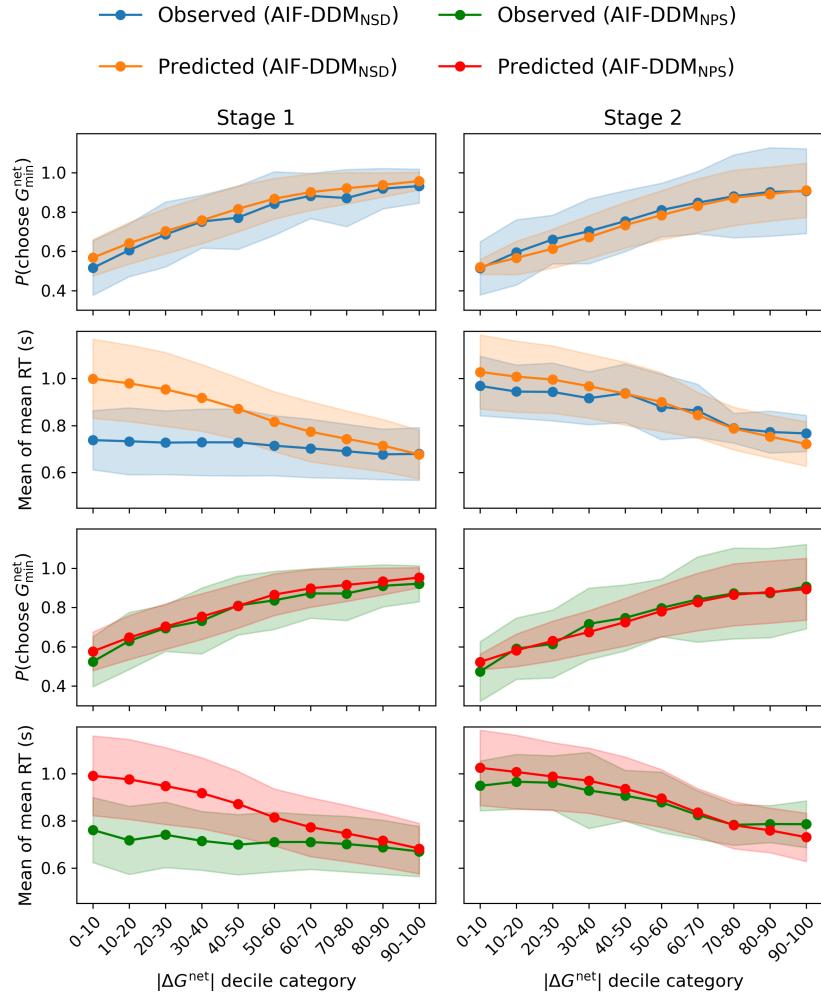
Fig. 6: Comparison of observed and model-predicted behaviour for AIF-DDM$_{\mathrm{NSD}}$ and AIF-DDM$_{\mathrm{NPS}}$ models.

G   Reaction time Q-Q analysis

To further investigate the reaction time (RT) misfit mentioned in Section 3.4, we generated Quantile-Quantile (Q-Q) plots comparing observed versus simulated within-participant RT deciles $(10^{th} - 90^{th})$. Simulated RT deciles were produced for each participant by running 50 simulated experiments with 1000 trials each (with the same reward volatility as the real experiment) using the best-fitted parameters. As shown in Fig. 7, the plots reveal that all models consistently overestimate the higher RT quantiles for both stages. This overestimation was expected, as the models' predictions are unconstrained, whereas the participants' observed RTs were truncated by a 2-second response deadline, with missed trials excluded from the analysis.
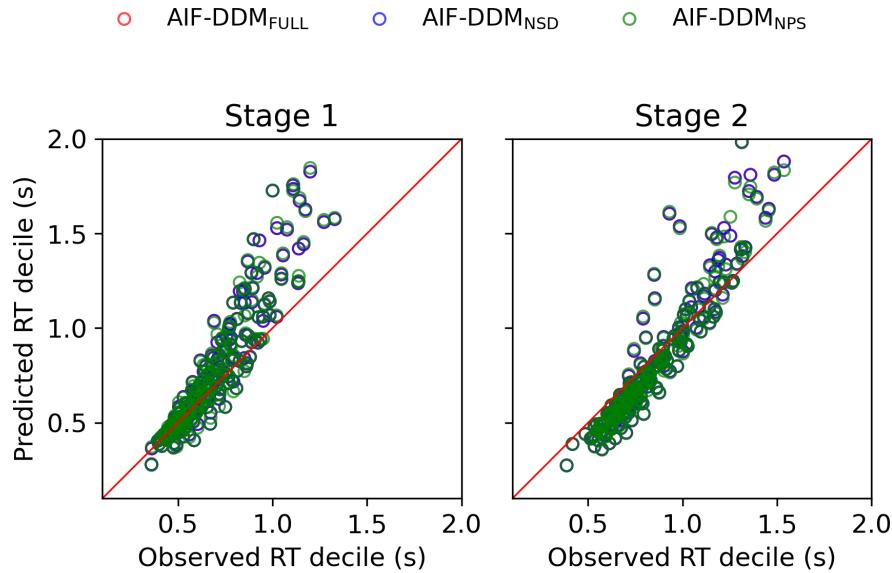

Fig. 7: Systematic overestimation of higher reaction time (RT) deciles by the AIF-DDM models. The Q-Q plots compare observed versus predicted RT deciles for Stage 1 (left) and Stage 2 (right). Points lying above the identity line (red) indicate that the model's predicted RT is higher than the observed RT. This effect is most pronounced for the longest reaction times, where predictions deviate furthest from the identity line. Note that the purple markers indicate an overlap of red and blue points.

acknowledge the authors of [17] for making their dataset publicly available and the authors of [4] for their open-source code.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Daunizeau, J., Adam, V., Rigoux, L.: Vba: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS computational biology **10**(1), e1003441 (2014)
2. Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J.: Model-based influences on humans' choices and striatal prediction errors. Neuron **69**(6), 1204–1215 (2011)
3. Fiedler, S., Glöckner, A.: The dynamics of decision making in risky choice: An eye-tracking analysis. Frontiers in psychology **3**, 335 (2012)
4. Gijsen, S., Grundei, M., Blankenburg, F.: Active inference and the two-step task. Scientific Reports **12**(1), 17682 (2022)
5. Hodson, R., Mehta, M., Smith, R.: The empirical status of predictive coding and active inference. Neuroscience & Biobehavioral Reviews **157**, 105473 (2024)
6. Lai, L., Gershman, S.J.: Human decision making balances reward maximization and policy compression. PLOS Computational Biology **20**(4), e1012057 (2024)
7. Lee, D.G., d'Alessandro, M., Iodice, P., Calluso, C., Rustichini, A., Pezzulo, G.: Risky decisions are influenced by individual attributes as a function of risk preference. Cognitive psychology **147**, 101614 (2023)
8. McDougle, S.D., Collins, A.G.: Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning. Psychonomic bulletin & review **28**(1), 20–39 (2021)
9. Myers, C.E., Interian, A., Moustafa, A.A.: A practical introduction to using the drift diffusion model of decision-making in cognitive psychology, neuroscience, and health sciences. Frontiers in Psychology **13**, 1039172 (2022)
10. Parr, T., Holmes, E., Friston, K.J., Pezzulo, G.: Cognitive effort and active inference. Neuropsychologia **184**, 108562 (2023)
11. Pedersen, M.L., Frank, M.J., Biele, G.: The drift diffusion model as the choice rule in reinforcement learning. Psychonomic bulletin & review **24**, 1234–1251 (2017)
12. Penny, W.D.: Comparing dynamic causal models using aic, bic and free energy. Neuroimage **59**(1), 319–330 (2012)
13. Proctor, R.W., Schneider, D.W.: Hick's law for choice reaction time: A review. Quarterly Journal of Experimental Psychology **71**(6), 1281–1299 (2018)
14. Rigoux, L., Stephan, K.E., Friston, K.J., Daunizeau, J.: Bayesian model selection for group studies—revisited. Neuroimage **84**, 971–985 (2014)
15. Schwartenbeck, P., Passecker, J., Hauser, T.U., FitzGerald, T.H., Kronbichler, M., Friston, K.J.: Computational mechanisms of curiosity and goal-directed exploration. elife **8**, e41703 (2019)
16. Shahar, N., Hauser, T.U., Moutoussis, M., Moran, R., Keramati, M., Consortium, N., Dolan, R.J.: Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. PLoS computational biology **15**(2), e1006803 (2019)

17. Feher da Silva, C., Hare, T.A.: Humans primarily use model-based inference in the two-stage task. Nature Human Behaviour **4**(10), 1053–1066 (2020)
18. Steinemann, N., Stine, G.M., Trautmann, E., Zylberberg, A., Wolpert, D.M., Shadlen, M.N.: Direct observation of the neural computations underlying a single decision. Elife **12**, RP90859 (2024)
19. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.: Scipy 1.0: fundamental algorithms for scientific computing in python. Nature methods **17**(3), 261–272 (2020)
20. Voskuilen, C., Ratcliff, R., Smith, P.L.: Comparing fixed and collapsing boundary versions of the diffusion model. Journal of mathematical psychology **73**, 59–79 (2016)
21. Wellcome Centre for Human Neuroimaging: Satistical parametric mapping software package. https://www.fil.ion.ucl.ac.uk (1991), accessed: 2025-5-10
22. Wiecki, T.V., Sofer, I., Frank, M.J.: Hddm: Hierarchical bayesian estimation of the drift-diffusion model in python. Frontiers in neuroinformatics **7**, 14 (2013)
23. Wilson, R.C., Collins, A.G.: Ten simple rules for the computational modeling of behavioral data. Elife **8**, e49547 (2019)
24. Yao, Z.F., Yang, M.H., Hsieh, S.: Neural correlates of span capacity during visual discrimination under varying cognitive demands. Scientific Reports **15**(1), 31071 (2025)
25. Zenon, A., Solopchuk, O., Pezzulo, G.: An information-theoretic perspective on the costs of cognition. Neuropsychologia **123**, 5–18 (2019)