Decision heuristics in contexts exploiting intrinsic skill

Neil M. Dundon[1,2], Jaron T. Colas[1], Neil Garrett[5], Viktoriya Babenko[1], Elizabeth Rizor[1], Dengxian Yang[3], Máirtín MacNamara[4], Linda Petzold[3], Scott T. Grafton[1]

[1]Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA 93106, USA
[2]Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University of Freiburg, 79104 Freiburg, Germany
[3]Department of Computer Science, University of California, Santa Barbara, CA 93106, USA
[4]Janssen Pharmaceutica NV, Beerse, Belgium
[5]School of Psychology, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK

Abstract

Heuristics can inform human decision making in complex environments through a reduction of computational requirements and a robustness to overparameterisation. However, tasks capturing the efficiency of reduced decision dimensionality typically ignore action proficiency in determining rewards. The value of movement parameterisation in sensorimotor control questions whether heuristics preserve efficiency when actions are non-trivial. We developed a novel selection-execution task requiring joint optimisation of action selection and spatio-temporal skill. Optimal choices could be determined by either a spatio-temporal forward simulation or a simpler spatial heuristic. Sequential-sampling models of action-selection response times parsimoniously distinguished human participants who adopted either strategy. Heuristics preserved broad decisional advantages over forward simulations. In addition, heuristics aligned with greater action proficiency, though predominantly through the core feature (spatial) shaping their decision policy. We accordingly reveal evidence that the dimensionality of information guiding action selection might be yoked to the granularity of plasticity in the motor system.

Introduction

In naturalistic settings, our cognitive architecture for making goal-oriented decisions typically resolves an ecological utility problem, integrating both extrinsic and intrinsic dynamics. Extrinsically, selected actions should maximise reward capture in line with a complex external state - a soccer player in possession of the ball must select the most rewarding action (shoot or pass) by incorporating such parameters as their location relative to the goalposts, availability of teammates, wind direction, readiness of the opposition goalkeeper, and so on. While the player might base their decision on exhaustive forward simulations across all possible actions, such a highly dimensional external state likely favours some manner of goal-oriented heuristic, i.e., a decision formed from a subset of all available external state information (e.g., if within 10 metres of the goalposts, shoot). Behavioural evidence verifies that a human decision policy can span different levels of planning complexity, with emerging neural evidence further suggesting that the brain harbours separate neural controllers for heuristics[1].

The logic underscoring heuristic adoption is at least two-fold. Heuristics first offer a trade-off between accuracy and available resources. That is, where exhaustive forward simulations might exceed computational resources or decision deadlines, heuristics offer a less laborious means to achieve a proxy for optimal action-selection policy[1,2]. An alternative "less-is-more" rationale, inspired by machine learning principles, considers heuristics as the optimal means to avoid overfitting in uncertain environments. That is, in uncertain environments, an overparameterised forward simulation will likely pick up on stochastic noise and create more prediction errors across choices than a function that uses fewer parameters, even if the latter function produces a biased estimate[3].

However, much like the areas of reinforcement learning and value-based decision making, evidence that humans adopt goal-oriented heuristics has emerged predominantly in contexts that do not consider intrinsic cost as a determining factor in reward yields. For example, in recent work[1], simple button presses in a virtual task emulated foraging outcomes that probabilistically imparted a positive (partial increase), negative (partial decrease) or nonlinearly negative (complete erasure) impact on ongoing reward scores; human participants adopted a heuristic stimulus-driven policy that primarily avoided the nonlinear outcome, consistent with accuracy-resource trade-offs. Meanwhile, the less-is-more principle has been empirically supported in forecasting contexts such as weather[3], investments[4] and sporting events[5]. Simple-action probabilistic emulations and forecasting can innovatively replicate much of the extrinsic reward-oriented cognitive challenges presented by dynamic naturalistic environments, however, they probe only one side of the ecological utility dilemma. Lost in both paradigm formats are additional cost dimensions associated with effort[6,7], motor plasticity[8,9], and a broader sense of agency[10], all of which integrate with external factors in the ultimate utility of selected actions in naturalistic settings[11,12].

To our knowledge, no study has characterised heuristic adoption by humans when they select state-appropriate actions in selection-execution contexts, i.e., not only is there a correct action for a given state, but the proficiency of a selected action subsequently scales the level of reward and generates independent intrinsic error distributions such as spatial and temporal motor skill. According to sensorimotor control theory, such intrinsic error distributions are often attenuated by increasing the parameterisation of movement, e.g., by implementing forward-models or simulations[27,28,29]. This raises the question: do heuristics preserve their efficiency when actions are non-trivial; or do action values derived from parameterised utility assessments justify preserved use of forward simulations? We additionally do not know how decision heuristics relate to phenotypic variation in skill. On the one hand, higher skill should improve both the time to generate, and the subsequent predictiveness of, forward simulations, potentially rendering parameterisation a more rewarding option for higher-skilled individuals. However, an alternative prediction stems from the computational underpinnings of how motor learning evolves. Here, the commonly held view is that model-based deliberation dominates early in learning[13], presumably while skill levels are also at their lowest. Thus if heuristics reflect a "model-free" antipode to model-based deliberation[14], we might expect them to characterise planning in systems that have reached greater proficiency with the plant's output.

Given their dynamic, non-punctuative nature, naturalistic states likely require action-selection deliberation to be a gradual process of evidence accumulation toward an action-deterministic criterion. Indeed, for decades, such a sequential-sampling framework has guided joint modeling of choice and the dynamics constraining its underlying response-time distribution, revealing comprehensive accounts of decision formation in perceptual contexts[15-18]. A specific class of sequential sampling, the drift-diffusion model (DDM), has more recently linked model-based values to choice dynamics in reinforcement learning[19,20] and value-based decision making[21,22]. The DDM approach has a potential benefit over discrete (e.g., logistic) models, stemming not just from comprehensiveness[23], but also from increased reliability[19] and robustness[18]. In the present work we use a DDM framework for the first time with data from a selection-execution task to distinguish people on their use of forward simulations or heuristics when selecting actions, and further hypothesise that these two groups might differ in terms of how skillfully they perform them.

The novel task we developed scales trialwise reward based on the joint optimisation of action selection (between two possible actions) and subsequent action execution. Action execution is indexed by a tractable reward function and action proficiency is also decomposable into spatial and temporal error dimensions. Meanwhile, action selection can be guided by simulating reward from each action (i.e., incorporating both spatial and temporal dynamics), or by a simpler spatial heuristic. The forward and heuristic strategies therefore differ by a single, identifiable degree of freedom (temporal dynamics). In addition, for each trial one possible action is low-cost, while the other is high-cost. All environments can be solved with either action, however, certain environments favour the high-cost action in terms of reward yield. These asymmetric action costs first amplify the influence of effort, skill and agency. They further allow us to cull participants who do not behave in a goal-oriented manner, i.e., rarely (or never) selecting the high-cost, high-potential reward action. Thirdly, the degree of parametric bias toward the low-cost action offers a novel complement to the commonly used approach of tracking intrinsic motor error as a means to characterise low skill systems.

Combining our novel task and seminal DDM framework, we first confirm that human participants can be identified by whether they predominantly use forward simulations or heuristics, with between-group parametric differences consistent with forward simulations requiring slower evidence accumulation to more conservative decision criteria. We next uncover strong evidence that heuristics remain efficient in action-execution contexts; heuristic planners made faster decisions, reached choice optimality sooner and, overall, obtained more reward. In addition, heuristic planners showed a striking skill profile. Overall, they demonstrated a higher level of spatial skill, relative to participants using forward simulations. However, unlike the latter group, heuristic participants demonstrated no learning across the task in the temporal domain. Their skill learning therefore centred on the key dimension incorporated into their planning strategy. Combined, our findings help unpack the broader dynamics of goal-oriented behaviour by revealing the first evidence of heuristic efficiency in a selection-execution context and that a yoked dimensionality might exist between planning and motor learning.

Results

Fifty-three healthy human participants performed 360 trials (six runs of 60) of our novel "boatdock" task (Figure 1), in which reward yields require joint optimisation of action selection and action execution. On each trial, participants select one of two cursors to pilot between a randomly drawn start-goal pairing (SG; Figure 1a). Each cursor accelerates continuously in three unique orthogonal directions (Figure 1b), burning fuel any time an accelerator button (throttle -

Figure 1c) is down. One cursor imparts a higher motor execution cost via an incongruent key-mapping (Figure 1c). However, trialwise reward is contingent on fuel conservation, such that a selection policy that harnesses the relative orthogonality of the two cursors' directionality, and selects the cursor better suited to each SG, will yield higher reward. The two cursors accelerate with the same nonlinear function, and deplete fuel with the same linear function, i.e., faster displacement is more fuel efficient (Figure 1d). A maximum docking displacement rule (Figure 1d), imposing a speed limit on arrival, imparts additional temporal control demands. Thus, in addition to fewer direction changes (spatial error), greater temporal control maximises reward (Figs. 1f, 1g). Finally, participants receive no reward for "catastrophic errors" (Figure 1e): when they run out of fuel, leave the grid, or attempt to dock above the maximum docking displacement.

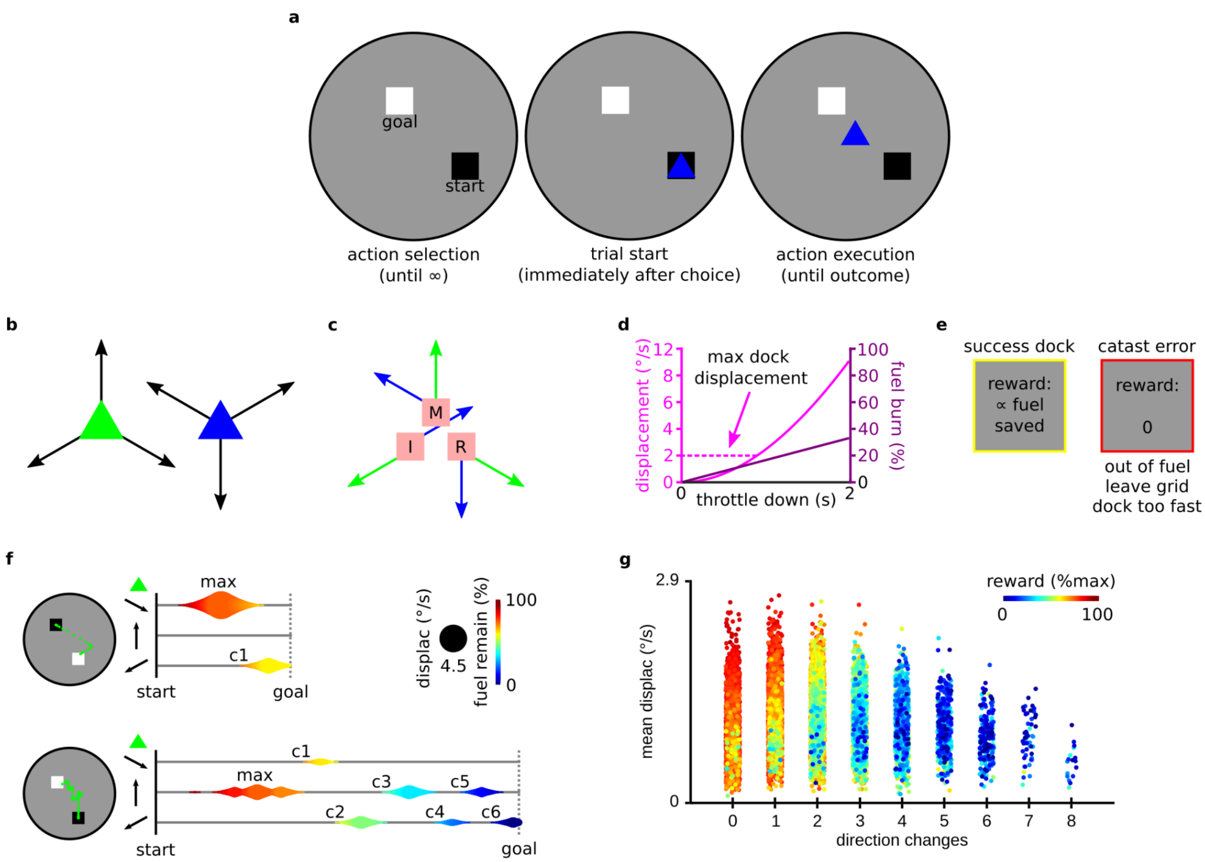

Figure 1 - Task outline. a, on each trial, participants pilot one of two cursors from a start to a goal (SG). **b,** each cursor can accelerate in three unique orthogonal directions. **c,** position of index (I), middle (M) and ring (R) finger of right hand on throttle buttons throughout the experiment, and cursor-specific throttle-vector mapping. Fuel burns any time a throttle is pushed down. Each trial allows six cumulative seconds of throttling before fuel depletes. **d**, throttle time linearly burns fuel, but nonlinearly increases displacement. Faster displacement is therefore more fuel efficient, however, a maximum dock displacement imparts additional temporal control requirements. **e,** successful docks yield a reward contingent on fuel conservation. This requires jointly maximising cursor choice for a given SG (action

selection) in addition to spatial and temporal precision (action execution). Trials containing catastrophic errors - running out of fuel, leaving the grid, or docking above maximum displacement - yield no reward. **f,** schematic of two similar SGs with the same cursor but different performance dynamics. Three horizontal lines in each panel chart activity over time separately for each vector, while each vortex relates to a single throttle pulse. Top panel utilises fewer direction changes (marked with c1,...,cn), reaches a higher maximum displacement (depicted by diameter of largest vortex) and yields higher reward (depicted by colour). **g**, reward (depicted by colour), yielded on every successful trial across all participants (individual markers), is a joint function of spatial and temporal precision.

Forward simulations vs heuristics

We hypothesised that participants might select the cursor yielding the highest reward from forward modeling of simulated routes, i.e., incorporating both the spatial and temporal constraints on reward optimisation into their choice (route planning; Figure 2a). Alternatively, participants might use a simple rule: select the cursor with a vector subtending the smallest angular offset to that of the SG, i.e., incorporating only spatial constraints on reward optimisation into their choice (heuristic; Figure 2b). Differences between these two approaches result in an imperfect correlation of action value across all trials. Specifically, the nonlinear nature of the temporal dynamics uniquely incorporated by route planning creates greater action-value deviance on trials where the SG covers smaller Euclidean distances (Figure 2d). Across all participants, the reaction time (RT) for action selection was slower in SGs where either strategy computed equivalent value for either cursor (Figure 2c), suggesting first that these policies modulate decision formation, and secondly, that difficulty or conflict (known to modulate drift parameter (μ) in perceptual contexts) could be further characterised in a DDM model.

DDM framework distinguishes individual-participant strategy

The DDM (Figure 2e) describes a noisy sequential sampling process, which originates at a starting point (b_cB), and accumulates evidence at an average "drift" rate (μ) before reaching a decision criterion or "boundary" (B or $-B$ for congruent or incongruent cursors, respectively). In perceptual contexts, difficulty reduces the gradient of evidence accumulation, which can be verified computationally when models containing two drift rates (e.g., μ_1 and μ_2) mapping respectively onto decisions presenting a high or low degree of difficulty, provide better model fits (Figure 2e). The goal of our DDM framework was to distinguish people based on whether they used a forward simulation or heuristic to guide decisions (Figs. 2a-b). We therefore formally considered a participant to be using a specific strategy if their evidence accumulation was best modulated by difficulty arising from it. To each participant's set of trialwise choices and RTs, we fitted a total of three DDMs, each containing free parameters for b_cB, B, and nondecision time t_0. The null model was constrained such that μ_1=μ_2=0, while the route-planner and heuristic models had two free parameters, μ_1 and μ_2, mapping respectively onto high or low difficulty as calculated by either forward simulations or angular offset (Figs. 2a-b.). Based on model fits[24] after adjusting for model complexity according to the Akaike information criterion with correction for finite sample size (AICc)[25,26], we confirmed that 19 participants' choice and RT data were best fitted by the route-planner model, 14 participants' data were best fitted by the heuristic model, while a third group of 20 participants were best fitted by the null model, indicating neither strategy-specific difficulty modulated the rate of their evidence accumulation (Table 1). We hence refer to these three groups respectively as the "route", "heuristic" and "nonplanner" groups (see Table 1 for group-specific DDM parameter and behaviour summaries). The remaining portion of these results will describe hypothesis-driven contrasts between the route and heuristic groups, in terms of DDM parameters, choice behaviour and skill. However, for completeness, we include data from the nonplanner group in Figure 3 and Supplementary Table 1, and also include a supplementary section summarising their parametric and skill findings (see Supplementary Materials: *Nonplanner group*).

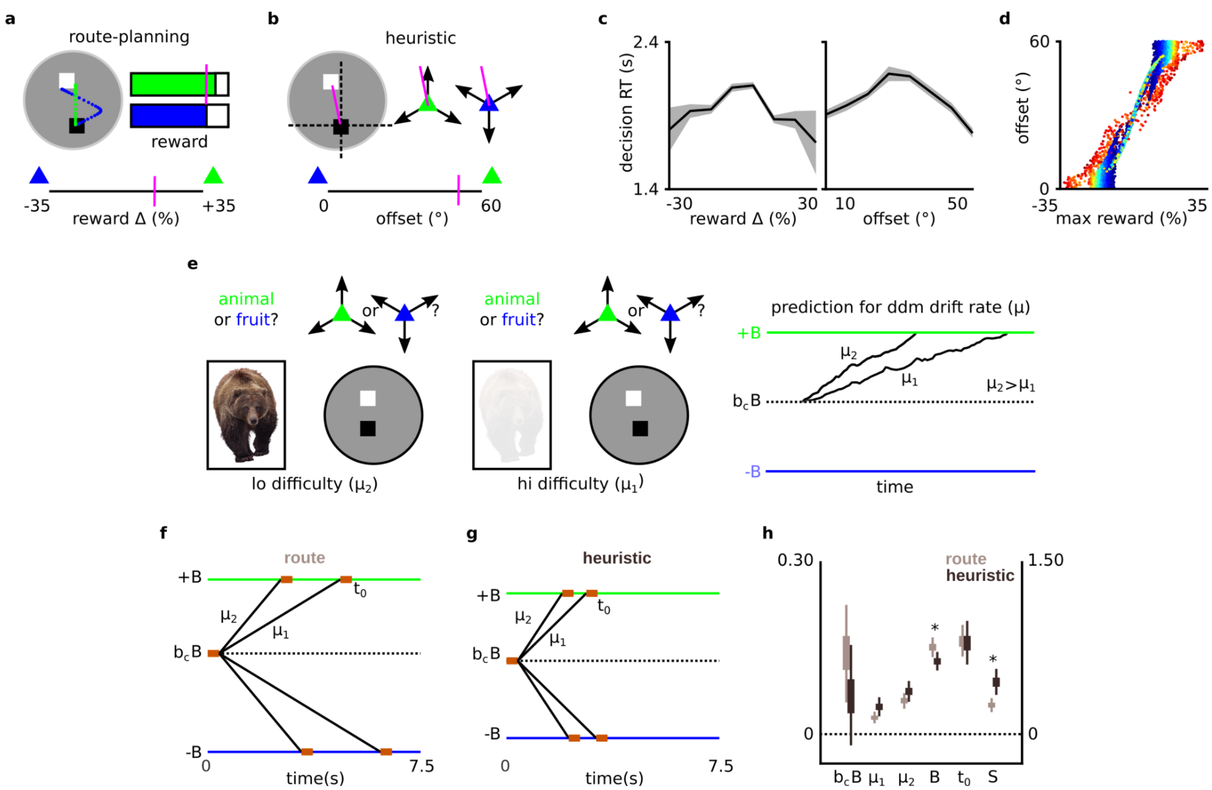

Figure 2 - Action-selection policy identified by sequential sampling framework. a-b, route planning selects the cursor in accordance with the delta between reward yields estimated from forward simulations with both cursors, i.e., incorporating both spatial and temporal task constraints into choice. A lower-dimensional heuristic instead selects the cursor with a displacement vector with the least angular offset to that described by the SG, i.e., incorporating only spatial information. **c**, across participants, both strategies create more difficult decisions, indexed by greater reaction time (RT) for action selection, on SGs where strategies ascribe equivalent value to both cursors. Here, solid black lines in each panel connect the means of nine RT bins after first sorting all participants' trials by relevant policy value. Gray shaded area depicts the standard error of the mean in each RT bin. **d**, strategy-specific action values are imperfectly positively correlated. Each individual marker describes the relation between individual action values derived from the route-planning and heuristic strategy across all trials from all participants. Hotter colors describe greater Euclidean distance between S and G, i.e., route-planning and heuristic strategy maximally deviates with shorter-distance

SGs. **e**, DDM framework, in which a noisy evidence accumulation process terminates at a decision criterion. We hypothesised that difficulty arising in our task would modulate the rate of evidence accumulation (drift rate μ) in a manner similar to that observed in perceptual tasks, and that comparison of strategy-specific difficulty modulation would identify individual-participant policy. For each participant's choice and RT data, two target models allowed separate drift rates μ_1 and μ_2 for high and low difficulty, respectively as per route-planning and heuristic strategy. Three groups of participants emerged, based on whether their data were best fitted by the route-planning (n=19), heuristic (n=14) or null (n=20) models (see also Table 1). **f-g**, scaled schematic of DDM profile estimated for route-planning (route) and heuristic groups. **h**, comparison of DDM parameters between route and heuristic group consistent with the former integrating additional information (i.e., temporal dynamics) into decision formation. Boxes and thin lines respectively represent the interquartile range (IQR) and highest density interval (HDI) of the posterior mean constraining individual-participant estimates of each parameter. Route group demonstrated a slower process of evidence accumulation toward a larger decision criterion; combining drift rates and decision criterion as part of a sensitivity metric $S=(\mu_1+\mu_2)/(2B(1+|b_cB|))$ we observe strong Bayesian evidence of group-level difference. This effect was primarily driven by the decision criterion, as the route group boundary (B) was also credibly higher than that of the heuristic group. In addition, the route group starting point (b_cB) was credibly above 0, indicating a bias toward the low-cost cursor. Asterisk indicates no overlap in groups' posterior HDIs for a given parameter. All parameters expressed in units of μ, except t_0 (in seconds). Note that t_0 and S parameters are aligned with the right axis for clarity.

Table 1. Participant groups and DDM parameters

	non	route	heuristic
n (site1:site2)	20 (7:13)	19 (6:13)	14 (3:11)
p(optimal cursor)	[0.477,0.500]	[0.668,0.690]	[0.706,0.730]
p(congruent cursor)	[0.716,0.736]	[0.564,0.587]	[0.528,0.555]
p(catastrophic error)	[0.169,0.186]	[0.120,0.135]	[0.110,0.126]
median response time (s)	[0.779,1.351]	[1.873,2.672]	[1.485,2.191]
sensitivity - S	0	[0.192,0.316]	[0.330,0.564]
drift rate, high difficulty - μ_1	0	[0.019,0.040]	[0.030,0.062]
drift rate, low difficulty- μ_2	0	[0.044,0.073]	[0.054,0.093]

boundary - B	[0.065,0.146]	[0.134,0.168]	[0.108,0.143]
starting point - b_cB	[0.127,0.405]	[0.059,0.236]	[-0.024,0.164]
nondecision time - t_0 (s)	[0.380,0.661]	[0.666,0.961]	[0.585,0.977]
route-planning DDM model: residual deviance - D_2	0.2	213.3	202.3
heuristic DDM model: residual deviance - D_2	0.1	171.2	238.8

Notes: values in square brackets for p(congruent cursor), p(optimal cursor) and p(catastrophic error) describe the bounds [upper,lower] of the HDI of group-level probability ($\theta_{v,g}$: see methods) from a summary binomial model. Values in square brackets for all other parameters describe HDI for group-level mean ($M_{v,g}$: see methods) from a summary Gaussian model. Drift rates (μ_1 and μ_2) describe rates of evidence accumulation per second toward Boundary (B), which is expressed in units of drift rate (μ). Starting point b_CB is expressed as a proportion from -B to B. Sensitivity, $S=(\mu_1+\mu_2)/(2B(1+|b_cB|))$. Higher values of residual deviances (D2) reflect better model fits across a group's participants relative to the null model.

We next tested whether individual-participant DDM parameters were consistent with group classifications ascribed by the model fit scores to the route and heuristic groups (Table 1, Figure 2h). By definition, route planning incorporates a larger volume of information into decisions, relative to the heuristic, which can be computationally indexed by more gradual evidence accumulation and or a broader decision criterion. We employed a Bayesian framework (see Methods) that estimated group-specific summaries of each DDM parameter listed in Table 1 in a single model, and further only considered strong Bayesian evidence of between-group parameter differences, i.e, where the highest density interval (HDI) of deterministic distributions of parameter differences did not contain zero. We first observed strong evidence of group difference for the sensitivity metric $S=(\mu_1+\mu_2)/(2B(1+|b_cB|))$ that combines the rate of evidence accumulation with the extent of the decision criterion (S Δ(route-heuristic) HDI=[-0.322,-0.056]). Credibly lower sensitivity amongst the route group is consistent with the above prediction that their employed policy integrates a greater volume of information, as this metric is low when decision formation is jointly constrained by a low rate of evidence accumulation and larger decision criterion (Figure 2f). This effect was primarily driven by the decision criterion, as we additionally observed the route group to have a credibly higher boundary (B) than the heuristic group (B Δ(route-heuristic) HDI=[0.002,0.050]), while not credibly differing across the two drift rates (μ_1 Δ(route-heuristic) HDI=[-0.036,0.002]; μ_2 Δ(route-heuristic) HDI=[-0.040,0.008]). Also of note, we observed a credible bias toward the low cost action amongst the the route (b_cB_{route} HDI=[0.057,0.233]), but not amongst the heuristic group ($b_cB_{\text{heuristic}}$ HDI=[-0.026,0.161]). Together, these findings provide parametric plausibility to the DDM classifications, and argue that the route group's lengthier trialwise decision deliberations stemmed from a greater volume of information integration, potentially stemming from a stronger bias toward using the low-cost cursor.

We next tested whether, on a broader macroscopic level, the route planner's computationally-intensive action selection delayed the emergence of optimality in their choice policy. For this, we summarised choice optimality across time-on-task using a hierarchical binomial model. The model estimated parameters of hierarchical Beta posterior distributions that constrained individual-participant binomial posteriors summarising their likelihood of selecting the optimal cursor in a run of trials, across a two-dimensional space described by group (route, heuristic, nonplanner) and run (1-6). From hierarchical Beta posteriors we deterministically computed credible ranges of group-specific choice optimality (θ) for each run, presented in Figure 3a. Consistent with optimality delay, the route group's choice behaviour (i.e., p(optimal)) did not credibly depart chance (0.50) until the fourth run of trials ($\theta_{route,run4}$ HDI=[0.533,0.882]), while in contrast, the heuristic group demonstrated above-chance choice optimality by the second run of trials ($\theta_{heuristic,run2}$ HDI=[0.501,0.898]; Figure 3a, see Supplementary Table 1 for each group-by-run θ HDI). This relative delay-to-optimality of approximately 120 trials provides further support that the route group mediated over a larger volume of evidence prior to action selection and further suggests a trade-off between how quickly a policy produces state-relevant choices, and the dimensionality of constituent planning.

Finally, in a supplementary analysis (see: Supplementary Materials - *Hierarchical logistic choice model*), we confirm that route and heuristic groups uniquely integrate extrinsic state information into choices (relative to nonplanners), and also confirm that the route group had a more pronounced bias toward the low-cost action, relative to the heuristic group, mirroring their bias revealed by the DDM. This supplementary analysis further revealed speculative evidence that the route group's planning strategy was not born purely out of risk-aversion, and that they instead potentially reserved high-cost action usage early in the task to states with longer SGs, where, due to nonlinear temporal task dynamics, state-appropriate choice offered disproportionately greater action value.

Comparisons of skill between route and heuristic groups

We have so far verified that a DDM framework parsimoniously distinguishes people on their likely use of forward simulations or heuristics during action selection in a selection-execution task, in a manner that is both parametrically consistent with the underlying characteristics of each strategy and in line with trade-offs between the expediency and profundity of policy formation. We next tested whether these two groups (identified solely using action-selection RT data) also differed in terms of action-execution skill and skill learning. We again employed a Bayesian framework that minimised the total number of fitted models and only considered strong evidence.

We first enumerated performance on each trial in terms of three skill variables: reward, spatial action execution and temporal action execution. Reward was the proportion of the fuel tank conserved on each trial, i.e., higher values reflect better performance on this measure which is modulated by the complete set of action-execution variables. Spatial action execution was the number of direction changes on each trial, i.e., lower values reflect better performance on this measure which is modulated specifically by spatial precision. Temporal action execution was the normalised difference between the cursor's maximum and final velocity, i.e., higher values reflect better performance on this measure that indexes proficiency in temporal task demands requiring high max-velocities for more fuel-efficient displacement, while arriving at the goal below the maximum threshold (Figure 1d), and ideally lower, to further preserve fuel. We next summarised performance in these three variables across the task using three separate hierarchical Bayesian models. Each model estimated hierarchical Gaussian (reward, temporal skill) or Poisson (spatial skill) posterior distributions that constrained individual-participant posterior distributions. We further fitted each model's hierarchical structure across a mixed three-dimensional space, i.e., first between-group (route, heuristic, nonplanner) and then within-group, separately for each run (1-6) and selected cursor (congruent, incongruent). In other words, each model estimated the credible ranges of group-mean performance for its given variable, separately for each run, and separately again for each cursor. Figures 3b-d and Supplementary Table 1 contain each group-by-cursor-by-run mu HDI for each measure and deterministic HDIs collapsed across run.

In terms of overall performance, the heuristic group garnered higher reward yields with the high-cost (incongruent) cursor (Figure 3b), which were at least partially attributable to a higher level of spatial skill (Figure 3c). Merging posteriors across runs, the route and heuristic groups showed no credible differences in reward yielded using the congruent cursor (mu Δ(route-heuristic) HDI=[-0.017,0.025], Figure 3b), however the heuristic group amassed credibly higher yields using the incongruent cursor (mu Δ(route-heuristic) HDI=[-0.055,-0.006], Figure 3b). This result was mirrored in spatial skill, where we again observed no credible between-group difference with the congruent cursor (mu Δ(route-heuristic) HDI=[-0.159,0.142], Figure 3c), but credibly fewer direction changes amongst the heuristic group with the incongruent cursor (mu Δ(route-heuristic) HDI=[0.011,0.358], Figure 3c). In contrast, we observed no between-group differences in temporal skill, with either the congruent (mu Δ(route-heuristic) HDI=[-0.018,0.056], Figure 3d) or incongruent (mu Δ(route-heuristic) HDI=[-0.068,0.007], Figure 3d) cursor. Given that the heuristic group reached state optimal choices more quickly (Figure 3a; Supplementary Table 1), i.e., they performed a higher volume of trials where their cursor selection theoretically reduced the need for direction changes, we re-ran the model with trialwise direction changes adjusted by the optimal solution for the cursor selected for each given trial (i.e., observed changes - ideal changes). This model (see: Supplementary Materials - *Hierarchical Poisson with choice-normalised spatial skill*) returned identical results, confirming that notwithstanding their better choices, the heuristic group independently demonstrated greater spatial precision while piloting the incongruent cursor.

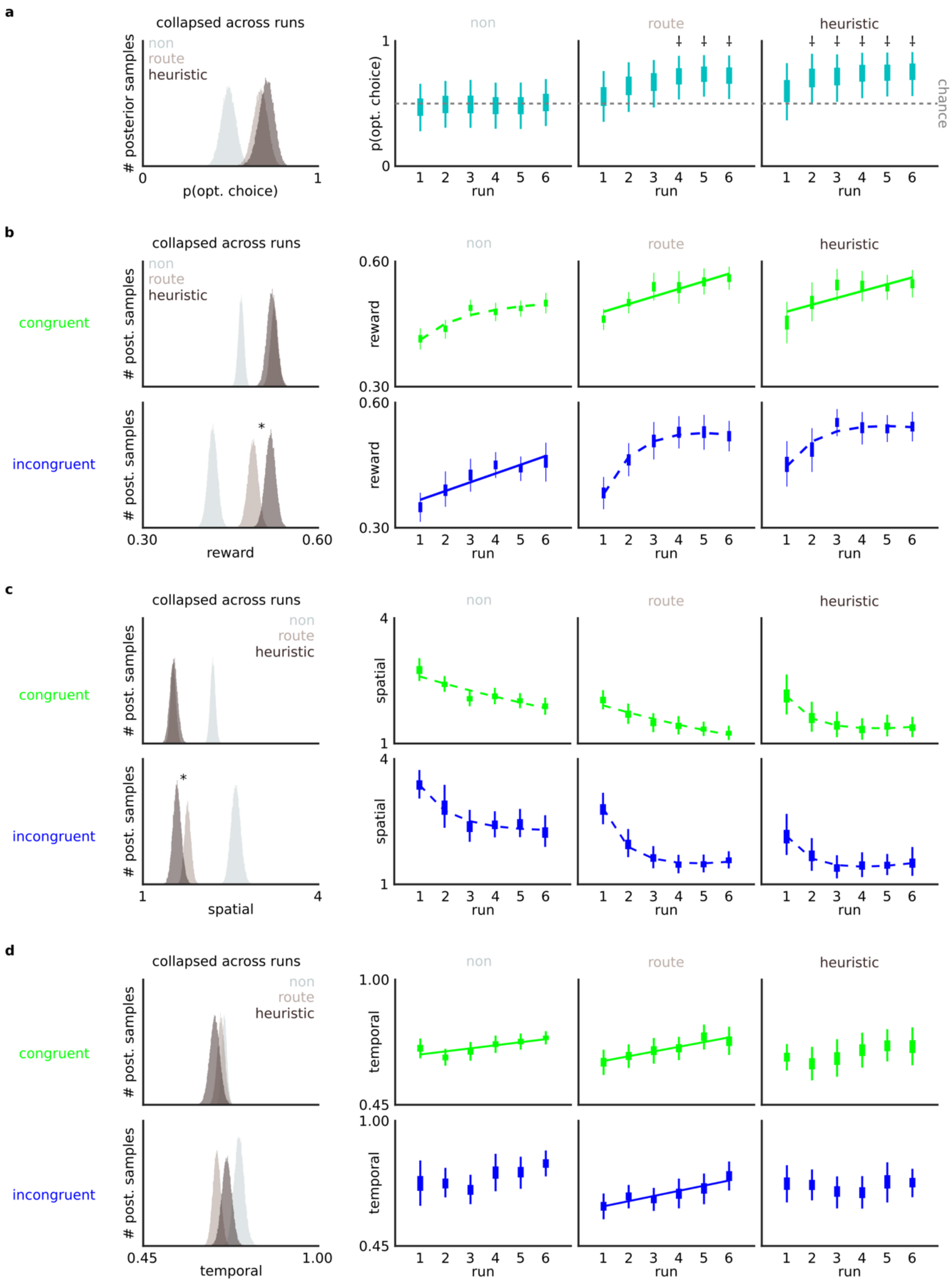

Figure 3 - Heuristics group reach choice optimality more quickly and show a spatial-specific skill advantage. a, consistent with classic decision-heuristic models, low dimensional planning aligns with faster trajectories toward choice

optimality. Hierarchical binomial model of choice behaviour demonstrates trade-off between the expediency and profundity of policy formation; heuristic group exceeded chance optimality by run 2, earlier than route group (run 4). ǂ reflects runs where HDI of group-level θ posterior did not subtend 0.50, i.e., where group-level proportion of choices were credibly above chance optimality. **b-d** skill and skill learning consistent with yoked dimensionality between planning and motor plasticity. Collapsing group-level posterior means across runs (skill) heuristic group yielded more reward with the high-cost cursor (**b**, histograms bottom panel), driven by superior spatial skill (**c**, histograms bottom panel), with no route-heuristic difference in temporal skill (**d**, histograms). Asterisk relates to credible difference between route and heuristic groups, i.e., that the HDI of the deterministic distribution of their difference (route - heuristic) does not contain 0. Additionally, while route and heuristic group demonstrated skill learning in terms of reward and spatial skill (**b**-**c**, line plots), route group uniquely demonstrated learning in the temporal domain (**d**, line plots). Boxes and thin lines in line plots respectively represent IQR and HDI of hierarchical posterior means constraining individual-participant posteriors for a given measure, run and cursor. In both histograms and line plots, reward is the proportion of fuel preserved per trial (higher better), spatial is the number of direction changes (fewer better) and temporal is the distance-normalized difference between max and final velocity (higher better). Time-on-task (skill learning) effects estimated from deterministic regression models fitted across draws from each run's posterior; credible ($0 \notin$ coefficient HDI) effects depicted by either a dashed (logarithmic) or solid (linear) line. Absence of any line reflects non-credible effect.

We next probed whether the route and heuristic groups differed in terms of time-on-task trajectories of specific skill features (skill learning). For this, we drew samples from each skill variable's uncollapsed runwise posterior distributions and tested with a deterministic regression model whether performance in each group-cursor dyad evolved across runs in either a linear or logarithmic fashion (linear and nonlinear time-on-task effects). Here we again observed divergence between the route and heuristic groups, with the route group's skill improvements encompassing a broader range of motor control features (Figure 3b-d, see Supplementary Table 1 for each group-by-run skill measure mu HDI, and HDIs of time-on-task effects). Specifically, the route group showed either linear or logarithmic improvement with each cursor across all variables, while in contrast, the heuristic group only showed strong evidence of improvement for reward yields and spatial skill, i.e., they demonstrated no time-on-task improvements with either cursor in the realm of temporal skill. To further support this null result, we conducted a follow-up analysis. This nonparametric analysis used more liberal criteria to establish temporal time-on-task effects at the individual-participant-level, and a binomial design that could confirm different trajectories between groups. Using summary values from individual-level posteriors, we quantified the number of participants from each group that showed either logarithmic or linear temporal time-on-task improvement, separately for each cursor. Confirmed via Bayesian binomial contrast, a greater proportion of the route group improved in temporal skill relative to the heuristic group, both using the congruent boat (10/19 vs 1/14; θ Δ(route-heuristic) HDI=[0.145,0.647]) and using the incongruent boat (10/19 vs 0/14; θ Δ(route-heuristic) HDI=[0.068,0.600]). This additional analysis confirmed that the route and heuristic groups diverged in the feature of temporal skill learning. In addition, this group-distinguishing feature of skill-learning - temporal dynamics - was also the feature which separated the two groups' planning styles. In other words, by definition, the heuristic group did not incorporate temporal dynamics into their action selection, and uniquely showed no skill learning in this domain. Together with the above findings relating to their overall superiority in spatial skill, these findings suggest that a yoked dimensionality might exist between planning and skill. I.e., in complex states, the number of features relevant (or not) for action-selection policy may predict the number of features most likely to undergo learning (or not) during action execution.

Discussion

At least two schools of thought lend plausibility to the idea that humans might achieve optimal long-term ecological yields by basing goal-oriented decisions on subsets of information available in complex states. On the one hand, if either time or computational resources are restricted, humans might pragmatically trade off state-optimal parameterisation for reduced processing requirements (accuracy-resource trade-off). On the other hand, decisions informed by fewer dimensions are more robust to the influences of misleading stochastic noise (less-is-more principle). In either case, extant knowledge on decision heuristics stems predominantly from action-trivial tasks that obviate intrinsic motor proficiencies in determining choice outcomes. The present work directly addressed this shortcoming.

We developed a novel task requiring joint optimisation of action planning (selecting state-appropriate low-cost or high-cost cursors) and action execution (controlling cursors proficiently). Focusing first on action planning, cursor selection could be shaped by either exhaustive forward simulations, incorporating both spatial and temporal task dynamics, or a lower-dimension spatial heuristic strategy. Using a between-group DDM framework we successfully parsed a wide pool of human participants based on which strategy-specific planning difficulty best modulated their evidence-accumulation rates. Exploring the dynamics of each group's choice behaviour in greater detail, we revealed strong group-level Bayesian evidence supporting the model classifications. Participants allocated to the route group (forward simulations) were constrained by a higher decision criterion, consistent with the idea that they needed to incorporate a larger volume of information into their choices. This group also showed an enduring bias toward low-cost actions, revealed in both a DDM and logistic-choice context, and further demonstrated a more sluggish trajectory toward choice optimality.

We next juxtaposed these two groups, who were classified solely on the basis of action-selection RT, in the separate context of action execution. Specifically, we probed the relation between decision-heuristic adoption and intrinsic skill, measuring the latter in terms of reward yielded per trial, and in terms of additional independent skill dimensions of spatial and temporal precision. We again revealed strong Bayesian group differences; heuristic adoption aligned with higher reward yields underscored by better spatial skill with high-cost actions, even after correcting for the influence of selections in state. Together with the route group showing a parametric bias toward the low-cost cursor, we interpret the combined data across our task's action-execution contexts as unambiguously supporting heuristic adoption in higher-skilled systems.

These core findings extend the remit of the accuracy-resource trade-off and less-is-more models to include contexts involving non-trivial action. In a task requiring exquisite spatio-temporal control of selected actions, decision heuristics nonetheless aligned with swifter individual action selections, faster trajectories to choice optimality and, ultimately, better overall reward yields. The skill advantage in the heuristic group additionally dissociates our findings from predictions that integrate a compensatory role of heuristics with theories from the human motor literature. A core rationale of sensorimotor-based forward models of action selection is that efference copies and predicted sensorimotor costs provide improved efficiency and robustness in the face of noisy sensory-prediction errors[27,28,29]. A corollary is that a low-skill system will struggle to parameterise forward models, increasing computational requirements (accuracy-resource) and/or generating a high volume of online noise-driven corrective action (less-is-more)[30,31]. In either case, a heuristic-as-compensation model would predict that the low skill system is better served by adopting a simpler heuristic decision strategy.

We observe that in a selection-execution context humans might not employ heuristics to compensate for higher intrinsic motor noise. Instead, planning dimensionality might align with progress along motor-learning trajectories previously observed in forced-choice contexts, (i.e., no selection required)[13,14]. Here, early in the acquisition of a novel motor skill, internal models that simulate action outcomes can expedite learning in exchange for high computational cost[13]. As participants then amass a wider cache of state transitions and successful experiences, control shifts from deliberative model-based planning to less taxing draws of state-appropriate motor outputs from memory[13]. Our findings are consistent with action selection following a similar qualitative trajectory at the cross-phenotypic level; participants with superior skill, i.e., farther along motor-learning trajectories, also used a less taxing policy to select actions.

In computational terms, the core difference between our task and paradigms previously exploring heuristics is the source (internal vs external) of its generative model. Trial outcomes in our task were determined solely by a joint function that integrated participants' cursor selection and its subsequent execution. In other words, outcome variance was fully determined by parameters (decision and performance) generated intrinsically by participants. In contrast, forecasting and computerised emulations typically employ extrinsic generative models, where outcome variance is a function of parameters beyond participants' control. Recent evidence from bandit tasks (a computerised emulation with an extrinsic generative model) further suggests that humans might overparameterise their choices when extrinsic forces determine their fate, resulting in apparently irrational summary behaviour such as probability matching[32]. However, probability matching dissipates as a function of increased agency, for example, with increased motor involvement in choice execution[32]. While it is premature to conclude that increased agency will globally drive the adoption of heuristics, our findings nonetheless predict that a low-skill system (i.e., low agency) will more likely overparameterise choices in a selection-execution context, than revert to heuristics. This inverse-agency-parameterisation framework is also consistent with emerging associations in clinical computational work, where sequelae such as overthinking (in anxiety[33]) and rumination (in depression[34]) align with excess deliberative model-based learning[35].

We reveal additional evidence that planning dimensionality and skill might not simply evolve independently along separate strands of a learning manifold. In our task, common kinematic variables parameterised both action value and the motor proficiency of subsequent execution. We were therefore able to probe how the depth of planning qualitatively aligned with the motor dimensions shaping both skill state and skill learning. As mentioned above, in terms of skill state, we first observed that the heuristic group's skill advantage was localised to the spatial dimension, with no overall group differences apparent in the temporal dimension. The heuristic group was therefore more skilled solely in the core feature of their decision policy. In a series of time-on-task analyses (skill learning) we additionally observed that the route group, employing the more granular planning, demonstrated skill learning across a broad array of motor-control features, including learning in temporal task dynamics. The heuristic group, in contrast, only showed skill learning in either the spatial or overall reward realms, i.e., no skill learning in the temporal domain (corroborated in a follow-up nonparametric analysis). Of note, a third nonplanner group, who never incorporated any state parameters into choice, and largely exploited the low-cost action, nonetheless improved across dimensions of motor-skill, including temporal skill (albeit only with the congruent cursor). In other words, the only group not showing credible plasticity in any temporal indices of skill learning was the heuristic group, i.e., the group who singularly used the spatial dimension of information when selecting actions.

These combined findings support the idea that a yoked dimensionality might exist between plans governing the selection of actions and the skill shaping their subsequent execution. In terms of a bottom-up framework, the spatial dominance in the heuristic group's planning-policy and skill advantage suggests that such dimensional yoking may be modulated by skill-first credit assignment[36]. In other words, higher execution proficiency stemming predominantly from spatial precision may have overweighted this dimension during planning. Previous research has indeed shown that human choice policy can be separately influenced by distinct dimensions of error depending on the reliability of their signals[37,38,39], and that increased agency might determine whether policy integrates either motor or reward-based errors[10]. A latter top-down framework is also supported by the apparent absence of temporal learning in the heuristic group. Note that the route and heuristic groups did not differ in terms of overall temporal skill, just that the heuristic group uniquely showed no time-on-task evolution in this domain. An intriguing implication of this pattern of results is that a controller that localises a cardinal subset of information for making state-appropriate action selections might itself be able to influence controllers of what it considers superfluous features of sensorimotor error.

Future behavioural enquiry into heuristics could employ advancements on our selection-execution framework to investigate the yoked dimensionality hypothesis and investigate its potential bottom-up and top-down underpinnings in more detail. An additional key outstanding question relates to the robustness of heuristic adoption over time. Given the tendency for learning-related configurations in the human brain to vary more across phenotypes than at the intra-subject level[40], we employed a between-groups analytic approach inspired by an increasing body of work that uses behavioural variance across the phenotype to increase robustness and reliability of hypothesis-specific brain activity[24,41]. While our DDM model parsimoniously distinguished human participants based on planning dimensionality, for power reasons, parameter estimations utilized all trials performed by participants. Our data therefore cannot inform any within-subject hypotheses regarding heuristic adoption; whether, for example, the route group would eventually reduce planning dimensionality with increased time-on-task. Though our logistic models revealed the route group's bias toward the low-cost cursor endured in later runs, suggesting their planning strategy may have held firm across the experiment, we cannot confirm whether they demonstrated a robust phenotypic trait or a relatively slower evolution along a trajectory of policy formation mutually traversed by both them and the heuristic group.

Conclusion

The association between decision heuristics and intrinsic skill has evaded description due to the arbitrary nature of action in computerised goal-oriented tasks. Here we used a novel task emulating both the decisional and motoric demands of goal-oriented behaviour in a dynamic environment. DDM models parsimoniously identified human participants who adopted heuristics, and later modeling unambiguously aligned this lower-dimensional planning strategy with higher skill, consistent with an inverse-agency-parameterisation model. We additionally observe that phenotypic variance in the intricacy of planning potentially maps onto the granularity of improvement in motor ability. Advancements in the behavioural assays of actions selected and executed will hopefully uncover the underlying causality, learning dynamics and neural underpinnings giving rise to this possible yoked dimensionality.

Materials and Methods

Participants and overview

We report data from a multi-site experiment, with 53 right-handed human participants recruited in total, via both word-of-mouth and the online participant-recruitment portal at the University of California, Santa Barbara (UCSB). 34 participants reported as female and the group had an average (standard deviation) age of 21.9 (3.05) years. Participants performed the experiment either in a behavioural-testing suite (site 1, n=16) or an fMRI context (site 2, n=37). We report only behavioural data in the present paper from both groups. Visual angle subtended by stimuli was constant for the two testing sites and neither site differed in terms of eventual DDM group classifications (see Supplementary Materials: *Site-specific DDM group classifications*). Participant remuneration was \$10 (\$20, site 2) per hour baseline rate, with an additional \$10 (\$20, site 2) contingent on performance. Testing at both sites took place during a single session. The Institutional Review Board at UCSB approved all procedures. Prior to participating, participants provided informed written consent. All stimuli were presented using freely available functions[42,43] written in MATLAB code, and unless otherwise stated all analyses were also conducted using custom MATLAB scripts.

Action selection-execution task: boatdock

Paradigm

Our task was a continuous, nonlinear adaptation of the discrete grid-sail task[13], extended such that reward yields require joint optimisation of action selection and action execution. All visual stimuli appear on a screen with a gray background ($RGB_{[0,1]}$=[0.500,0.500,0.500]). In each trial (Figure 1a), they select one of two cursors, depicted by equilateral triangles (side length=0.830 °), to pilot from a start (S) to a goal (G), respectively depicted by a black ($RGB_{[0,1]}$=[0,0,0]) and white ($RGB_{[0,1]}$=[1,1,1]) square (side length=1.37 °). The SG pair appears within a circular grid (radius=3.82 °) centered on the screen center. Locations of the SG are drawn with uniform probability on each trial, constrained such that neither element falls within 0.320 ° of the grid perimeter, and their centres are at least 0.957 ° apart.

Each cursor displaces in three deterministic directions (Figure 1b.), mapping onto the same three separate response buttons ("throttles") operated by the right hand for the duration of the experiment (Figure 1c). One "congruent" cursor displaces at angles $7\pi/6$ (index finger), $\pi/2$ (middle finger), and $11\pi/6$ (ring finger) in a reference frame where $\pi/2$ aligns with the vertical meridian of the screen (Figure 1c). The other "incongruent" cursor displaces at angles $5\pi/6$, $\pi/6$ and $3\pi/2$, via one of two sets of spatially incongruent throttle-mappings, selected with uniform (p=0.500) probability for each subject (an example mapping is in Figure 1c). For the entire experiment, the congruent and incongruent cursors are identified by a different color, green $RGB_{[0,1]}$=[0,1,0] and blue $RGB_{[0,1]}$=[0,0,1], determined with uniform (p=0.500) probability before each participant's session.

For every frame a single throttle is down, the cursor will accelerate in that direction (see Supplementary Materials for specific Acceleration dynamics) and a one unit of fuel is also subtracted from an allocation of 360 units provided for each trial. Participants therefore have a total 6 s throttle time on each trial before fuel depletes (refresh rate=60 Hz). Following a successful "dock" (see below) a screen informs participants of the fuel conserved, expressed as a proportion of the starting tank. No other exogenous cue is provided to participants regarding the size of the initial fuel allocation, or its rate of depletion.

Trial structure

Each trial initiates with the action-selection period, signified by the appearance of an SG pair within a grid ("action selection", Figure 1a). Participants have no time limit to select their desired cursor with the middle or index finger of their left hand, respectively using "a" or "z" of a standard keyboard (site 1) or buttons 1 and 2 (i.e., the two most leftward) of a six-button bimanual response box[44] (site 2). Finger-cursor mapping (i.e., index→congruent, middle→incongruent, or vice versa) is determined every twenty trials by uniform (p=0.500) probability, prompted throughout the action-selection period by a silhouette of a hand (9.49 °-by-9.49 °) below the grid, with the relevant cursor above the relevant finger. Once an action is selected, the action-execution period immediately begins, signified by the silhouette prompt disappearing and the selected cursor spawning at the centre of S ("trial start", Figure 1a). Participants now pilot the cursor from S to G with their right hand, using the "v" (index), "h" (middle) or "m" (ring) buttons on the keyboard (site 1) or buttons 4-6 on the right side of the response box (site 2). Action execution lasts until one of four possible trial outcomes. A successful "dock" is achieved if the cursor enters a 0.479 °-radius circular threshold (not visible to participants) centred on the centre of G, at a velocity no greater than 1.920 °/s. Alternatively, three catastrophic errors can occur if participants (i) run out of fuel, i.e., cumulative throttle time greater than 6 s; (ii) leave the grid; or (iii) enter the circular G threshold at a velocity greater than 1.920 °/s. Once a trial outcome is achieved, a feedback screen immediately informs participants of the outcome, respectively, "WELL DONE!", "OUT OF GAS!","LEFT THE GRID!" or "TOO FAST!", presented at the centre of the screen along with "SCORE: $", where $ is either the proportion of fuel preserved (for successful docks) or 0 for all catastrophic errors. The feedback remains on the screen for 1 s, followed by a blank grey inter-trial-interval screen lasting one, two or three seconds (determined on each trial with uniform probability p=0.333). Participants performed 360 choice trials in total, portioned into six runs of 60 trials. Interlaced between choice runs were 20 practice trials, on which scores do not count toward the final bonus, forcing ten trials with both the congruent and incongruent cursor in pseudorandom order.

Dependent variables

To enumerate action values derived from route planning we first computed forward simulations of the optimal routes on each (simulation procedure described in Supplementary materials). We subtract the total frames spent accelerating during the optimal route (λ) from the starting fuel bank of 360 units to estimate the maximum reward obtainable on a given trial. To enumerate action values derived from a simpler spatial heuristic we computed the angles (in °) between the vector of a trial's SG and each vector on the incongruent cursor. The vector creating the smallest angle (which we term the "offset") quantifies action value from this heuristic on a raw scale where values close to 0 reflect an SG perfectly aligning with one of the incongruent vectors and values close to 60 reflect an SG perfectly aligning with one of the congruent vectors.

We enumerated reaction time (RT) for action selection as the time elapsed between the time of the first frame of the action-selection screen (described above) and the time of cursor selection. We coded optimal selection as incongruent cursor on trials with offset<30 and congruent cursor on trials with offset>30 (optimal cursor selection did not differ depending on which action value (route vs heuristic) is computed).

We enumerated skill performance on each trial in terms of reward, spatial action execution and temporal action execution. Reward was the amount of fuel conserved. All modeling of reward used raw units (i.e., on a scale of 0 to 360) to allow Gaussian likelihood functions, however for clarity in reported results we present findings as a proportion of the tank preserved (from 0 to 1). Spatial action execution was the number of direction changes, i.e., a count of how many times a different throttle was pressed relative to the one previous. Temporal action execution was the difference between the cursor's maximum velocity recorded during action-execution (in °/s), and the final velocity (in °/s) taken at the moment the cursor crossed the circular threshold around G, normalised by the distance covered by the SG (in °).

Data analysis

Computational modeling

We modeled action planning leading up to cursor selection with variants of a standard drift-diffusion model[45,46,47]. The full models included five free parameters: high-difficulty drift rate μ_1, low-difficulty drift rate μ_2, boundary B, starting point $b_C B$, and nondecision time t_0. The boundaries for congruent and incongruent choices were defined as B and $-B$, respectively. Hence a positive $b_C B$ relates to a congruency bias. Parameters were necessarily constrained as follows: $0 \leq \mu_1 \leq \mu_2$, $\mu_2 \geq 0$, $B > 0$, $-1 < b_C B < 1$, and $t_0 > 0$. Noise was represented as the standard deviation of diffusion with a fixed scaling parameter σ=0.1.

We compared three types of models: two route-planning models (with one or two drift rates), two heuristic models (with one or two drift rates), and the null (i.e., nonplanning) model. For route-planning models, we determined difficulty by dividing trialwise differences in reward yields (between the simulated optimal routes for either cursor) into five bins. For the heuristic models, we determined difficulty by dividing trialwise offsets into five bins. The five difficulty bins corresponded to drift rates of $-\mu_2$, $-\mu_1$, 0, μ_1, and μ_2. We constrained single-drift-rate models such that μ_1=μ_2 to minimise penalties for additional degrees of freedom, and the null model such that μ_1=μ_2=0 to represent insensitivity to the onscreen information.

We fitted candidate models to empirical distributions of choices and RTs at the level of individual subjects using maximum-likelihood estimation and the chi-square fitting method[48]. We calculated the frequencies of either choice and the 10, 30, 50, 70, and 90% quantiles (i.e., six bins) of their respective RT distributions for each difficulty level. We optimised free parameters with respect to overall goodness of fit for given subjects using iterations of the Nelder-Mead simplex algorithm with randomised seeding[49]. We adjusted for model complexity when comparing models that differed in degrees of freedom using the Akaike information criterion with correction for finite sample size (AICc)[25,26].

Three participant groups were defined by the results of model fitting following penalisation. The "route" and "heuristic" groups included those who were best fitted by a route-planning or heuristic model, respectively, according to the AICc. For assignment to the "nonplanner" group, adding free parameters for planning would not yield a significant improvement in goodness of fit relative to the null model without sensitivity to either route or heuristic information.

Bayesian models

We sampled all Bayesian posterior distributions using No U-Turn sampling (NUTS) Hamiltonian Monte Carlo, implemented with the PyMC3 package[50] in custom Python scripts. Unless otherwise specified, each model's posterior distributions were sampled across four chains of 10000 samples (40000 total), with an additional initial 10000 samples per chain (40000 total) discarded after tuning the sampler's step-size to an acceptance threshold of 0.95 (80000 samples combined), with further convergence criteria that no chains contain any divergences and no posterior's $\hat{R}$ value, estimating the ratio of variance within the n=4 chains to the variance of the pooled chains, greater than 1 (see:[51]). Unless otherwise stated, dependent variables were z-score normalised across participants prior to fits. We calculated minimum-width Bayesian credible intervals of relevant posteriors from their chains, using the default settings for Highest Density Interval (HDI) calculation in the arviz package[52].

A pair of models first estimated summaries of group-specific behaviour (reported in Table 1). A single Gaussian model first summarised continuous variables, accounting for eight variables in total. First, the four variables applicable to each group identified by the DDM framework, specifically: starting point - b_cB, boundary - B, and nondecision time - t_0, in addition to median RT. In addition, the three variables applicable only to the route and heuristic groups, specifically: drift rate, hi difficulty - μ_1, drift rate, lo difficulty- μ_2 and sensitivity - S. This model assumed individual participant (n) values (y) for each variable (v) were characterised by a separate Gaussian likelihood function, further depending on n's group-allocation (g(n): route, heuristic or nonplanner), i.e., $y_{n,v} \sim \mathcal{N}(M_{v,g(n)}, \Sigma_{v,g(n)})$. Each variable was z-score normalised separately (but across all subjects) prior to fitting, and we respectively assigned each $M_{v,g(n)}$ and $\Sigma_{v,g(n)}$ an uninformed Gaussian and half-Gaussian prior: $M_{v,g(n)} \sim \mathcal{N}(\mu=0, \sigma=10)$ and $\Sigma_{v,g(n)} \sim \text{half}\mathcal{N}(\sigma=10)$. Three separate binomial models then estimated summaries of behaviour as measured by three binomial variables that applied to all groups: p(congruent cursor), p(optimal cursor) and p(catastrophic error). For the n(g) participants in each group (g), each summary model used a Binomial likelihood function $y_g \sim \text{Bin}(\theta_g, t_g)$, where y_g and t_g are n(g)-element vectors, respectively enumerating the number of observed instances reported by each individual participant in a group (y_g) and their total number of trials (t_g). In each model, we assigned each θ_g an uninformed prior from the beta distribution: $\theta_g \sim \text{Beta}(\alpha=1, \beta=1)$.

We used a hierarchical Bayesian binomial model to estimate the credible ranges of group-specific choice optimality (p(optimal cursor)), separately for each run. The hierarchical structure used Binomial likelihood functions to summarise the number of optimal cursor selections (y) made by each participant (n) for all trials (t) in a given run (r), $y_{n,r} \sim \mathrm{Bin}(\theta_{n,r}, t_{n,r})$. The model constrained $\theta_{n,r}$ posteriors with separate hierarchical group (g(n)) and run-specific Beta distributions, i.e.: $\theta_{n,r} \sim \mathrm{Beta}(\alpha_{g(n),r}, \beta_{g(n),r})$. Each $\alpha_{g(n),r}$ and $\beta_{g(n),r}$ were assigned uninformed priors from a half-Student's T distribution, i.e.: $\alpha_{g(n),r} \sim \mathrm{HalfStudentT}(\sigma=10, \nu=10)$ and $\beta_{g(n),r} \sim \mathrm{HalfStudentT}(\sigma=10, \nu=10)$, bounded to never draw values of $\alpha_{g(n),r}=0$ or $\beta_{g(n),r}=0$. Run-specific group-level deterministic posterior estimates of optimal choice ($\theta_{g(n),r}$) were calculated by drawing 10,000 independent samples (k) from relevant $\alpha_{g(n),r}$ and $\beta_{g(n),r}$ posteriors and computing the mean of the resulting kth Beta distribution, i.e., $\theta_{g(n),r,k} = \alpha_{g(n),r,k} / (\alpha_{g(n),r,k} + \beta_{g(n),r,k})$.

We used two separate hierarchical Bayesian Gaussian models to estimate the credible ranges of group-mean performance in the two continuous action-execution variables (reward and temporal skill), separately for each run, and separately again for each cursor. In each model, the hierarchical structure used Gaussian likelihood functions to summarise each (n) participant's trialwise measures across all trials in a given run (r), separately for each cursor (c), i.e.: $y_{n,r,c} \sim \mathcal{N}(\mu_{n,r,c}, \exp(\sigma_{n,r,c}))$. The model constrained $\mu_{n,r,c}$ and $\sigma_{n,r,c}$ posteriors with separate hierarchical group (g(n)), run (r) and choice-specific (c) Gaussian distributions, i.e.: $\mu_{n,r,c} \sim \mathcal{N}(M(\mu)_{g(n),r,c}, \Sigma(\mu)_{g(n),r,c})$ and $\sigma_{n,r,c} \sim \mathcal{N}(M(\sigma)_{g(n),r,c}, \Sigma(\sigma)_{g(n),r,c})$. Each $M(\mu)_{g(n),r,c}$ and $M(\sigma)_{g(n),r,c}$ were assigned uninformed Gaussian priors ($\sim\mathcal{N}(\mu=0, \sigma=10)$), while each $\Sigma(\mu)_{g(n),r,c}$ and $\Sigma(\sigma)_{g(n),r,c}$ were assigned uninformed half-Gaussian priors ($\sim\mathrm{half}\mathcal{N}(\sigma=10)$). Note that the model for reward was fitted to a continuous measure, scoring fuel conserved on a scale of 0 to 360, but for clarity, we adjusted runwise and collapsed HDIs (division by 360), also prior to computing any HDIs related to between-comparisons, to express results as a proportion of fuel preserved. Time-on-task betas, however, relate to unadjusted posteriors.

We used a hierarchical Bayesian Poisson model to estimate the credible ranges of group-mean performance in spatial skill, separately for each run, and separately again for each cursor. In each model, the hierarchical structure used Poisson likelihood functions to summarise each (n) participant's trialwise direction changes across all trials in a given run (r), separately for each cursor (c), i.e.: $y_{n,r,c} \sim \mathrm{Pois}(\exp(\mu_{n,r,c}))$. The model constrained $\mu_{n,r,c}$ posteriors with separate hierarchical group (g(n)), run (r) and cursor-specific (c) Gaussian distributions, i.e.: $\mu_{n,r,c} \sim \mathcal{N}(M(\mu)_{g(n),r,c}, \Sigma(\mu)_{g(n),r,c})$. $M(\mu)_{g(n),r,c}$ and $\Sigma(\mu)_{g(n),r,c}$ were respectively assigned uninformed Gaussian ($\sim\mathcal{N}(\mu=0, \sigma=10)$) and half-Gaussian priors ($\sim\mathrm{half}\mathcal{N}(\sigma=10)$). For clarity in reported results, we re-adjusted runwise and collapsed HDIs (exponential transform), also prior to computing any HDIs related to between-comparisons, to discount the use of $\exp(\mu_{n,r,c})$ in the likelihood function. Time-on-task betas, however, relate to unadjusted posteriors.

For both the hierarchical Gaussian and Poisson skill models, separately for each group (g) and choice (c), we enumerated deterministic posteriors of overall skill level by averaging each posterior sample across runs, i.e., for each posterior sample, $M(\mu)_{g,c} = 1/6 \sum_{r=1}^{6} M(\mu) g, r, c$. We then enumerated deterministic linear and logarithmic time-on-task effects $b_{g,c}$ by drawing posterior samples from $M(\mu)_{g,r,c}$. Specifically, on each (k) of 40,000 draws, we computed the kth column of

$b_{g,c}$ ($b_{g,c,k}$), where $b_{g,c,k}=(X^TX)^{-1}X^TY_{g,c,k}$. Here, $Y_{g,c,k}$ is a six-element column vector containing an independent draw from each run (r) of $M(\mu)_{g,r,c}$ and matrix X is a three-column matrix respectively containing six constant terms (1), z-scored linear $x\in(1,2,...,6)$ and z-scored logarithmic $x\in(\ln(1),\ln(2),...,\ln(6))$ regressors. The second and third rows of resulting 3-by-40,000 matrix $b_{g,c}$ respectively contained deterministic posteriors for linear and logarithmic time-on-task effects. Where logarithmic time-on-task effects were credible ($0 \notin$ HDI), we considered that group-cursor time-on-task effect to be logarithmic even if a linear effect was also observed. Note, as specified above, that in reported results, we present the HDIs of time-on-task coefficients (linear and logarithmic) fitted to unadjusted runwise posteriors, i.e., before we made any adjustment to posteriors for intuitive presentation of runwise/collapsed HDIs.

For the individual-participant-level nonparametric analysis of temporal skill, we computed the median of each $\mu_{n,r,c}$ posterior from the relevant Gaussian skill model. Separately for each cursor we regressed the six-element vector of participant's run-specific median values, first as a function of an intercept and a linear time-on-task regressor (z-scored linear $x\in(1,2,...,6)$), and then as a function of an intercept and a z-scored logarithmic regressor $x\in(\ln(1),\ln(2),...,\ln(6))$. If either model's regressor (x) was significant (determined by 95% coefficient confidence intervals not containing 0), we considered that participant time-on-task$^+$ for that cursor and skill variable. We compared proportions of time-on-task$^+$ participants (y) between groups (g), separately for each cursor, by fitting Binomial likelihood function $y_g \sim \text{Binomial}(\theta_g,n_g)$, assigning each θ_g an uninformed prior from the beta distribution: $\theta_g \sim \text{Beta}(\alpha=1,\beta=1)$.

In all above cases, we consider strong evidence of credible effects as follows: for comparison of parameters to criterion values (e.g., a regression coefficient above 0, or a likelihood above 0.50, etc.) we required the entire HDI of that parameter to not include the criterion value. For comparison of two parameters we required the HDI of the deterministic distribution of their difference (posterior A - posterior B) to not contain 0. Note that two HDIs might overlap, but that this deterministic distribution of difference may yet still not contain 0.

Acknowledgements
The research was supported by award #W911NF-16-1-0474 from the Army Research Office and by the Institute for Collaborative Biotechnologies under Cooperative Agreement W911NF-19-2-0026 with the Army Research Office.

Competing interests
Authors declare no financial, and no non-financial, competing interests.

References

1. Korn, C. W. & Bach, D. R. Minimizing threat via heuristic and optimal policies recruits hippocampus and medial prefrontal cortex. *Nat. Hum. Behav.* **3**, 733–745 (2019).
2. Gilovich, T., Griffin, D., Kahneman, D. & Cambridge University Press. *Heuristics and Biases: The Psychology of Intuitive Judgment*. (Cambridge University Press, 2002).
3. Gigerenzer, G. & Brighton, H. Homo heuristicus: why biased minds make better inferences. Top. Cogn. Sci. 1, 107–143 (2009).
4. DeMiguel, V., Garlappi, L. & Uppal, R. Optimal versus Naive Diversification: How

Inefficient Is the 1/N Portfolio Strategy? *Heuristics*, 644–664 (2011).
5. Scheibehenne, B. & Bröder, A. Predicting Wimbledon 2005 Tennis Results by Mere Player Name Recognition. *Heuristics*, 613–623 (2011).
6. Richter, M., Gendolla, G. H., & Wright, R. A. Three Decades of Research on Motivational Intensity Theory: What We Have Learned About Effort and What We Still Don't Know. in *Advances in Motivation Science* **3**, 149–186 (2016).
7. Shadmehr, R., Huang, H. J. & Ahmed, A. A. A Representation of Effort in Decision-Making and Motor Control. *Curr. Biol.* **26**, 1929–1934 (2016).
8. Chapman, C. S. *et al.* Short-term motor plasticity revealed in a visuomotor decision-making task. *Behav. Brain Res.* **214**, 130–134 (2010).
9. Diquelou, M. C., Griffin, A. S. & Sol, D. The role of motor diversity in foraging innovations: a cross-species comparison in urban birds. *Behav. Ecol.* **27**, 584–591 (2016).
10. Parvin, D. E., McDougle, S. D., Taylor, J. A. & Ivry, R. B. Credit Assignment in a Motor Decision Making Task Is Influenced by Agency and Not Sensory Prediction Errors. *J. Neurosci.* **38**, 4521–4530 (2018).
11. Wolpert, D. M. & Landy, M. S. Motor control is decision-making. *Curr. Opin. Neurobiol.* **22,** 996-1003 (2012).
12. Haggard, P. Human volition: towards a neuroscience of will. *Nat. Rev. Neurosci.* **9**, 934–946 (2008).
13. Fermin, A. S. R. *et al.* Model-based action planning involves cortico-cerebellar and basal ganglia networks. *Sci. Rep.* **6**, 1–14 (2016).
14. Korn, C. W. & Bach, D. R. Heuristic and optimal policy computations in the human brain during sequential decision-making. *Nat. Commun.* **9**, 325 (2018).
15. Usher, M. & McClelland, J. L. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* **108**, 550–592 (2001).
16. Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).
17. Forstmann, B. U., Ratcliff, R. & Wagenmakers, E.-J. Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annu. Rev. Psychol.* **67**, 641–666 (2016).
18. Peters, J. & D'Esposito, M. The drift diffusion model as the choice rule in inter-temporal and risky choice: A case study in medial orbitofrontal cortex lesion patients and controls. *PLoS Comput. Biol.* **16**, e1007615 (2020).
19. Shahar, N. *et al.* Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. *PLoS Comput. Biol.* **15**, e1006803 (2019).
20. Ballard, I. C. & McClure, S. M. Joint modeling of reaction times and choice improves parameter identifiability in reinforcement learning models. *J. Neurosci. Methods* **317**, 37–44 (2019).
21. Colas, J. T. Value-based decision making via sequential sampling with hierarchical competition and attentional modulation. *PLoS One* **13**, e0203093 (2017).
22. Fontanesi, L., Gluth, S., Spektor, M. S. & Rieskamp, J. A reinforcement learning diffusion decision model for value-based decisions. *Psychon. Bull. Rev.* **26**, 1099–1121 (2019).
23. Pedersen, M. L., Frank, M. J. & Biele, G. The drift diffusion model as the choice rule in reinforcement learning. *Psychon. Bul. Rev.* **24**, 1234–1251 (2017).
24. Colas, J. T., Pauli, W. M., Larsen, T., Tyszka, J. M. & O'Doherty, J. P. Distinct prediction

errors in mesostriatal circuits of the human brain mediate learning about the values of both states and actions: evidence from high-resolution fMRI. *PLoS Comput. Biol.* **13**, e1005810 (2017).

25. Akaike, H. A new look at the statistical model identification. *IEEE T. Automat. Contr.* **19**, 716-723 (1974).
26. Hurvich, C. M. & Tsai, C.-L. Regression and time series model selection in small samples. Biometrika **76**, 297–307 (1989).
27. Miall, R. C. & Wolpert, D. M. Forward Models for Physiological Motor Control. *Neural Networks* **9**, 1265–1279 (1996).
28. Maruyama, S. *et al*. Cognitive control affects motor learning through local variations in GABA within the primary motor cortex. *Sci. Rep.* **11**, 18566 (2021).
29. Shadmehr, R., Smith, M. A. & Krakauer, J. W. Error correction, sensory prediction, and adaptation in motor control. *Annu. Rev. Neurosci.* **33**, 89–108 (2010).
30. Tunik, E., Houk, J. C. & Grafton, S. T. Basal ganglia contribution to the initiation of corrective submovements. *Neuroimage* **47**, 1757–1766 (2009).
31. Grafton, S. T. & Tunik, E. Human basal ganglia and the dynamic control of force during on-line corrections. *J. Neurosci.* **31**, 1600–1605 (2011).
32. Green, C. S., Benson, C., Kersten, D. & Schrater, P. Alterations in choice behavior by manipulations of world model. *P Natl Acad Sci USA*. **107**, 16401-16406 (2010).
33. Gagne, C., Dayan, P. & Bishop, S. J. When planning to survive goes wrong: predicting the future and replaying the past in anxiety and PTSD. *Curr. Opin. Behav. Sci.* **24**, 89–95 (2018).
34. Huys, Q. J. M., Daw, N. D. & Dayan, P. Depression: a decision-theoretic analysis. *Annu. Rev. Neurosci.* **38**, 1–23 (2015).
35. Hunter, L. E., Meer, E. A., Gillan, C. M., Hsu, M. & Daw, N. D. Increased and biased deliberation in social anxiety. *Nat Hum Behav.* **6**, 146–154 (2022).
36. McDougle, S. D. *et al*. Credit assignment in movement-dependent reinforcement learning. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6797–6802 (2016).
37. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
38. Lee, S. W., Shimojo, S. & O'Doherty, J. P. Neural computations underlying arbitration between model-based and model-free learning. *Neuron* **81**, 687–699 (2014). *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6797–6802 (2016).
39. O'Doherty, J. P. *et al*. Why and how the brain weights contributions from a mixture of experts. *Neurosci. Biobehav. Rev.* **123**, 14–23 (2021).
40. Bassett, D. S. *et al*. Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7641–7646 (2011).
41. Schönberg, T., Daw, N. D., Joel, D. & O'Doherty, J. P. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J. Neurosci.* **27**, 12860–12867 (2007).
42. Brainard, D. H. The Psychophysics Toolbox. *Spatial Vision* **10**, 433-436 (1997).
43. Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision **10**, 437-442 (1997).
44. Current Designs Incorporated. Philidelphia, PA.
45. Stone, M. Models for choice-reaction time. *Psychometrika* **25**, 251-260 (1960).
46. Laming, D. R. J. *Information theory of choice-reaction times*. (Academic Press, Oxford, United Kingdom. 1968).

47. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **85**, 59-108 (1978).
48. Ratcliff, R., & Tuerlinckx, F. Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychon. B. Rev.* **9**, 438-481 (2002).
49. Nelder, J. A., & Mead, R. A simplex method for function minimization. *Comput. J.* **7**, 308-313 (1965).
50. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* **2**, e55 (2016).
51. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P. C. Rank-normalization, folding, and localization: An improved $\widehat {R} $ for assessing convergence of MCMC. *Bayesian Analysis* **16**, 667–718 (2021).
52. Kumar, R., Carroll, C., Hartikainen, A. & Martin, O. ArviZ a unified library for exploratory analysis of Bayesian models in Python. *J. Open Source Softw.* **4**, 1143 (2019).

Supplementary materials

Nonplanner group

In addition to the 33 participants identified by our DDM framework as likely employing one of two planning strategies, 20 participants were best fitted by the null model (skill variables of this group are included in Supplementary Table 1). While our hypotheses principally focused on parametric and skill differences between the two classes of planners (route and heuristic), we briefly comment here on the nonplanner group. Of all groups, nonplanners showed the fastest overall decision time and strongest DDM bias toward the congruent cursor (Table 1), consistent with an action-selection policy that did not integrate the external state. However, despite demonstrating no evidence of state-appropriate action selection (Figure 3b.), largely stemming from an over-reliance on the congruent cursor (Table 1), nonplanners nonetheless exhibited skill learning during the execution portion of our task (Figure 3a). They improved with both cursors in terms of reward yield and spatial precision, but only demonstrated improved temporal dynamics with the congruent cursor, i.e., the cursor they exploited to yield reward.

Supplementary Table 1: group-by-run skill measure mu HDI, HDIs collapsed across runs and HDIs of time-on-task effects

cursor/variable	HDI	non	route	heur
p(optimal choice)	θ run 1	[0.279,0.658]	[0.354,0.760]	[0.366,0.823]
	θ run 2	[0.307,0.681]	[0.433,0.830]	[0.501,0.898]†
	θ run 3	[0.306,0.689]	[0.469,0.848]	[0.514,0.893]†
	θ run 4	[0.299,0.667]	[0.533,0.882]†	[0.543,0.906]†
	θ run 5	[0.297,0.668]	[0.555,0.887]†	[0.557,0.908]†
	θ run 6	[0.321,0.696]	[0.536,0.885]†	[0.560,0.938]†

cong/reward	mu run 1	[0.415,0.468]	[0.463,0.515]	[0.430,0.532]
	mu run 2	[0.442,0.488]	[0.504,0.557]	[0.484,0.580]
	mu run 3	[0.495,0.538]	[0.535,0.603]	[0.535,0.613]
	mu run 4	[0.484,0.530]	[0.528,0.608]	[0.533,0.608]
	mu run 5	[0.496,0.535]	[0.550,0.614]	[0.537,0.598]
	mu run 6	[0.504,0.554]	[0.562,0.619]	[0.543,0.610]
	mu (runs coll)	[0.487,0.505]	[0.542,0.567]	[0.533,0.566]
	β(run)	[0.091,0.191]*	[0.103,0.218]*	[0.058,0.230]*
	β(log(run))	[0.004,0.397]*	[-0.003,0.480]	[-0.004,0.714]
incong/reward	mu run 1	[0.338,0.410]	[0.369,0.449]	[0.425,0.537]
	mu run 2	[0.375,0.462]	[0.450,0.532]	[0.461,0.568]
	mu run 3	[0.412,0.493]	[0.491,0.584]	[0.549,0.615]
	mu run 4	[0.446,0.510]	[0.519,0.584]	[0.528,0.608]
	mu run 5	[0.439,0.499]	[0.519,0.598]	[0.534,0.600]
	mu run 6	[0.437,0.533]	[0.515,0.601]	[0.536,0.690]
	mu (runs coll)	[0.431,0.462]	[0.500,0.534]	[0.530,0.565]
	β(run)	[0.101,0.268]*	[0.146,0.308]*	[0.055,0.243]*
	β(log(run))	[-0.042,0.599]	[0.236,0.897]*	[0.008,0.763]*
cong/spatial	mu run 1	[2.497,3.043]	[1.817,2.277]	[1.696,2.651]
	mu run 2	[2.237,2.614]	[1.470,1.956]	[1.284,2.000]
	mu run 3	[1.891,2.264]	[1.271,1.728]	[1.151,1.766]
	mu run 4	[1.939,2.335]	[1.218,1.657]	[1.092,1.605]
	mu run 5	[1.844,2.208]	[1.178,1.523]	[1.171,1.690]
	mu run 6	[1.687,2.098]	[1.082,1.428]	[1.153,1.640]
	mu (runs coll)	[2.114,2.279]	[1.433,1.605]	[1.406,1.656]
	β(run)	[-0.156,-0.076]*	[-0.211,-0.105]*	[-0.212,-0.048]*
	β(log(run))	[-0.303,0.007]	[-0.378,0.054]	[-0.703,-0.024]*

incong/spatial	mu run 1	[3.053,3.736]	[2.438,3.190]	[1.702,2.686]
	mu run 2	[2.354,3.380]	[1.650,2.323]	[1.324,2.106]
	mu run 3	[2.018,2.782]	[1.384,1.916]	[1.143,1.696]
	mu run 4	[2.123,2.748]	[1.265,1.706]	[1.174,1.766]
	mu run 5	[2.130,2.804]	[1.287,1.713]	[1.220,1.726]
	mu run 6	[1.895,2.654]	[1.380,1.791]	[1.204,1.893]
	mu (runs coll)	[2.439,2.748]	[1.662,1.874]	[1.452,1.726]
	β(run)	[-0.180,-0.063]*	[-0.241,-0.128]*	[-0.195,-0.015]*
	β(log(run))	[-0.458,-0.008]*	[-0.728,-0.262]*	[-0.744,-0.035]*
cong/temporal	mu run 1	[0.650,0.736]	[0.579,0.688]	[0.598,0.712]
	mu run 2	[0.619,0.692]	[0.609,0.711]	[0.555,0.700]
	mu run 3	[0.640,0.721]	[0.628,0.738]	[0.571,0.733]
	mu run 4	[0.674,0.749]	[0.643,0.743]	[0.610,0.761]
	mu run 5	[0.688,0.757]	[0.659,0.793]	[0.637,0.776]
	mu run 6	[0.710,0.768]	[0.665,0.787]	[0.621,0.784]
	mu (runs coll)	[0.685,0.716]	[0.667,0.712]	[0.641,0.700]
	β(run)	[0.027 0.133]*	[0.043 0.205]*	[-0.018,0.187]
	β(log(run))	[-0.411,0.034]	[-0.250,0.385]	[-0.530,0.279]
incong/temporal	mu run 1	[0.623,0.819]	[0.565,0.676]	[0.638,0.798]
	mu run 2	[0.660,0.787]	[0.612,0.713]	[0.647,0.780]
	mu run 3	[0.630,0.757]	[0.601,0.701]	[0.612,0.750]
	mu run 4	[0.686,0.848]	[0.612,0.741]	[0.611,0.751]
	mu run 5	[0.697,0.835]	[0.630,0.763]	[0.638,0.811]
	mu run 6	[0.750,0.862]	[0.689,0.814]	[0.660,0.783]
	mu (runs coll)	[0.716,0.776]	[0.654,0.701]	[0.678,0.737]
	β(run)	[-0.004,0.219]	[0.048,0.217]*	[-0.096,0.114]
	β(log(run))	[-0.657,0.270]	[-0.389,0.256]	[-0.592,0.232]

*Notes: non=nonplanner; heur=heuristic; cong=congruent cursor; incong=incongruent cursor; spatial=spatial skill; temporal=temporal skill; coll.=collapsed across runs; runwise and collapsed HDIs for reward have been re-adjusted (division by 360) to express reward as a proportion of fuel preserved; *=time-on-task coefficient credibly dearts 0;+ proportion of choices credibly above chance optimality (0.50).*

Hierarchical logistic choice model

We used a hierarchical Bayesian logistic regression model to assess the group-specific modulation of choice (p(incongruent)) as a function of an intercept (β0), trial offsets (β1; i.e., the trialwise enumeration of heuristic value) and the Euclidean distance (in screen pixels) of trial SGs (β2). The hierarchical structure used Bernoulli likelihood functions to characterise choice likelihood for each individual participant (n) and trial (t), i.e.: $y_{n,t} \sim \text{Bernoulli}(p_{n,t})$, where $p_{n,t}$ is computed with a deterministic logistic transition function $S(x_{n,t})$, where $x_{n,t}=\beta0_n+\beta1_n*\text{offset}_{n,t}+\beta2_n*\text{distance}_{n,t}$. The model constrained coefficient posteriors fitted to each participant's set of trials with separate hierarchical group-specific (g(n)) Gaussian distributions, i.e.: $\beta0_n \sim \mathcal{N}(M(\beta0)_{g(n)},\Sigma(\beta0)_{g(n)})$, $\beta1_n \sim \mathcal{N}(M(\beta1)_{g(n)},\Sigma(\beta1)_{g(n)})$ and $\beta2_n \sim \mathcal{N}(M(\beta2)_{g(n)},\Sigma(\beta2)_{g(n)})$. Each $M(\beta0)_{g(n)}$, $M(\beta1)_{g(n)}$ and $M(\beta2)_{g(n)}$ were assigned uninformed Gaussian priors ($\sim\mathcal{N}(\mu=0,\sigma=10)$), while each $\Sigma(\beta0)_{g(n)}$, $\Sigma(\beta1)_{g(n)}$ and $\Sigma(\beta2)_{g(n)}$ were assigned uninformed half-Gaussian priors ($\sim\text{half}\mathcal{N}(\sigma=10)$). Both regressors were z-score normalised across all trials from all subjects prior to fitting. Finally, we fitted two iterations of this model, one using trials from the early phase of the task (first three runs), and a second using trials from the late phase of the task (final three runs).

Results of this hierarchical logistic regression model are summarised below in Supplementary Table 2. This model first bolstered the DDM by demonstrating the route and heuristic group uniquely integrated state information into action selection. During both early and late phases of the task, the route ($\beta1_{\text{route}}$mu HDI early=[-0.865,-0.483]; $\beta1_{\text{route}}$mu HDI late=[-1.559,-1.014]) and heuristic group ($\beta1_{\text{heuristic}}$mu HDI early=[-1.466,-0.601]; $\beta0_{\text{heuristic}}$mu HDI late=[-2.080,-1.201]), incorporated route offsets optimally into choice; note that their credibly negative coefficient HDIs reflect increased likelihood of selecting the incongruent cursor when offset angle was low, i.e., suited to the incongruent cursor (offset was normalised to vectors on the incongruent cursor; see: Methods - Dependent variables). In addition, this model supported the finding from the DDM relating to the route group's bias. The route group uniquely showed a bias to the congruent cursor in both early and late phases of the task, ($\beta0_{\text{route}}$mu HDI early=[-0.682,-0.240]; $\beta0_{\text{route}}$mu HDI late=[-0.423,-0.104]), which was not credibly evident in the heuristic group in either instance ($\beta0_{\text{heuristic}}$mu HDI early=[-0.480,0.011]; $\beta0_{\text{heuristic}}$mu HDI late=[-0.401,0.073]). No groups credibly modulated their choice by the distance covered by a route's start-goal pairing (SG), in either early or late phases of the task (all HDIs for β2 subtend 0 in Supplementary Table 2). However, of note, the trending positive distance parameter estimate for the route group in the early phase suggests first that their planning strategy was not born out of risk-aversion, (which instead would have been characterised by incongruent selection on shorter SGs). Though we can only speculate on a non-credible finding, if the route group selectively used the high-cost incongruent cursor early in primarily longer SGs, they may have been reserving its usage for situations where optimal choice was disproportionately beneficial, due to the nonlinear temporal task physics (Figure 1d, Results: *Forward simulations vs heuristics,* Figure 2d).

Supplementary Table 2: Hierarchical logistic model of choice behaviour parameters

task phase	parameter	non	route	heur
early	β0 (int.) mu	[-1.690,-0.780]*	[-0.682,-0.240]*	[-0.480,0.011]
late	β0 (int.) mu	[-1.471,-0.672]*	[-0.423,-0.104]*	[-0.401,0.073]
early	β1 (offset) mu	[0.016,0.196]*	[-0.865,-0.483]*	[-1.466,-0.601]*
late	β1 (offset) mu	[-0.059,0.172]	[-1.559,-1.014]*	[-2.080,-1.201]*
early	β2 (distance) mu	[-0.092,0.100]	[-0.015,0.136]	[-0.175,0.045]
late	β2 (distance) mu	[-0.127,0.071]	[-0.083,0.087]	[-0.112,0.123]

*Notes: non=nonplanner; heur=heuristic; int.=intercept; *=coefficient credibly dearts 0;*

Hierarchical Poisson model with choice-normalised spatial skill

To dissociate the heuristic group's superior spatial skill with the incongruent cursor from their overall more optimal choice behaviour, we used a hierarchical Bayesian Poisson model to estimate the credible ranges of group-mean performance in spatial skill, using a measure which had been normalised by the optimal number of direction changes in the simulated solution (see Supplementary Materials: *Optimal route simulations*). This normalisation took the number of direction changes made on each trial, and subtracted from that the number of direction changes made by the optimal solution for the specific cursor chosen on that trial (i.e., not necessarily normalised to the optimal cursor for a given route, but the selected cursor). Due to a small number of resulting trials (0.3%, across all subjects) containing a negative value (never lower than -1), we added a constant (1) to all trials, to ensure the lowest value was 0, suitable for a Poisson likelihood function. With this normalisation, higher values reflect worse spatial skill, i.e., more direction changes relative to cursor-optimal. As with the unnormalised model, we fitted the model separately for each run, and separately again for each cursor. In each model, the hierarchical structure used Poisson likelihood functions to summarise each (n) participant's trialwise direction changes across all trials in a given run (r), separately for each cursor (c), i.e.: $y_{n,r,c}$~Pois(exp($\mu_{n,r,c}$)). The model constrained $\mu_{n,r,c}$ posteriors with separate hierarchical group (g(n)), run (r) and cursor-specific (c) Gaussian distributions, i.e.: $\mu_{n,r,c} \sim \mathcal{N}(M(\mu)_{g(n),r,c},\Sigma(\mu)_{g(n),r,c})$. $M(\mu)_{g(n),r,c}$ and $\Sigma(\mu)_{g(n),r,c}$ were respectively assigned uninformed Gaussian ($\sim\mathcal{N}(\mu=0,\sigma=10)$) and half-Gaussian priors ($\sim$half$\mathcal{N}(\sigma=10)$). For clarity in reported results, we re-adjusted runwise and collapsed HDIs (exponential transform, followed by subtraction of -1), also prior to computing any HDIs related to between-comparisons, to discount first the use of exp($\mu_{n,r,c}$) in the likelihood function, and then the constant added to all trials prior to fitting. Time-on-task betas, however, relate to unadjusted posteriors.

Results of this model are summarised below in Supplementary Table 3. Crucially, collapsing across runs, we see the heuristic group demonstrating credibly fewer direction changes with the incongruent cursor (mu Δ(route-heuristic) HDI=[0.005,0.331]), supporting the interpretation of the finding from the main paper (see: Results - *Comparisons of skill between route and heuristic groups*) that their superior incongruent spatial skill is independent to the navigational consequences of their choices.

Supplementary Table 3: group-by-run spatial skill, normalised by cursor selection

cursor/skill	HDI	non	route	heur
cong/spatial	mu run 1	[1.664,2.216]	[1.085,1.522]	[1.000,1.886]
	mu run 2	[1.408,1.776]	[0.742,1.195]	[0.610,1.228]
	mu run 3	[1.092,1.430]	[0.555,0.972]	[0.481,1.046]
	mu run 4	[1.073,1.479]	[0.519,0.891]	[0.422,0.872]
	mu run 5	[0.982,1.323]	[0.464,0.758]	[0.477,0.889]
	mu run 6	[0.879,1.261]	[0.363,0.669]	[0.432,0.861]
	mu (runs coll.)	[1.276,1.436]	[0.707,0.861]	[0.704,0.926]
	β(run)	[-0.148,-0.076]*	[-0.173,-0.092]*	[-0.178,-0.052]*
	β(log(run))	[-0.282,0.000]	[-0.308,0.029]	[-0.546,-0.016]*
incong/spatial	mu run 1	[2.190,2.912]	[1.664,2.408]	[0.912,1.875]
	mu run 2	[1.568,2.593]	[0.943,1.578]	[0.629,1.347]
	mu run 3	[1.166,1.924]	[0.674,1.160]	[0.456,0.978]
	mu run 4	[1.266,1.915]	[0.640,1.018]	[0.534,1.100]
	mu run 5	[1.217,1.918]	[0.594,0.966]	[0.516,1.002]
	mu run 6	[1.032,1.782]	[0.611,0.994]	[0.489,1.113]
	mu (runs coll.)	[1.586,1.901]	[0.953,1.150]	[0.749,1.007]
	β(run)	[-0.178,-0.066]*	[-0.210,-0.115]*	[-0.157,-0.008]*
	β(log(run))	[-0.421,0.007]	[-0.573,-0.185]*	[-0.567,0.022]

*Notes: non=nonplanner; heur=heuristic; cong=congruent cursor; incong=incongruent cursor; spatial=cursor-normalised spatial skill; coll.=collapsed across runs; runwise and collapsed HDIs have been re-adjusted (subtraction of -1) to discount the constant added to all trials prior to fitting; *=time-on-task coefficient credibly departs 0.*

Site-specific DDM group classifications

To test group allocations from the DDM for each site, we fitted a summary Bayesian multinomial model. The model used a k=3 multinomial likelihood function to characterise the counts (#) for each group classification y=[#(route) #(heuristic) #(nonplanner)], separately for the participants (n_s) each site (s) y_s~Multinomial(Θ_s,n_s). We assigned Θ_s an uninformed prior from a Dirichlet distribution Θ_s~Dirichlet(α=[1,1,1]). Results from this model (summarised in Supplementary Table 3 below) confirmed the DDM ascribed similar group allocations for both testing sites.

Supplementary Table 4: site-specific DDM group allocations

parameter	HDI (site 1)	HDI (site 2)
Θ p(route)	[0.167,0.569]	[0.213,0.493]
Θ p(heuristic)	[0.054,0.386]	[0.164,0.435]
Θ p(nonplanner)	[0.220,0.633]	[0.215,0.492]

Acceleration dynamics
At a resolution of 60 Hz, cursor position during action execution is updated for each frame f by adding a two-element vector ($\mathbb{D}$) to the cursor's position at frame f-1. $\mathbb{D}$ is computed using $D = \sum_{v=1}^{3} \mathbb{P}_v * \mathbb{V}(v)$. Here, $\mathbb{P}_v$ is the two-element vector (x,y) in screen coordinates describing a Euclidean displacement of 0.320 ° in the direction of a given throttle (v). $\mathbb{V}$(v) scales each coordinate in $\mathbb{P}_v$ in accordance with nonlinear acceleration by using $\mathbb{V}(v) = f(\mathbb{T}(v))$. Where$f(x) = -0.011x^3 + 0.167x^2$. For every frame a given throttle (v) is down, the relevant element of three-element vector $\mathbb{T}$ (i.e., $\mathbb{T}$(v)) increases by 0.017 s, and for every frame a throttle is released, $\mathbb{T}$(v) decreases by 0.017 s until it reaches 0. Elements of $\mathbb{T}$ therefore update separately and gradually at this fixed rate, meaning non-zero momentum from one vector can continue influencing the displacement of the cursor after its release and while another throttle is down, allowing curvilinear two-dimensional displacement (see top panel of Figure 1f). However, if more than one throttle is down for a given frame, each element of $\mathbb{T}$ decreases by 0.017 s (unless already at 0), precluding participants from using simultaneous throttle pulsing to create additional displacement angles outside of the six afforded across the two cursors.

Optimal route simulations
To enumerate action values derived from route planning we first computed forward simulations of the optimal routes (i.e., with the highest reward yield) from S to G for each cursor on each trial. Separately for each cursor, we first assessed whether the SG on each trial afforded a single linear displacement with one of its vectors from S that would intersect the circular threshold around G (point of intersection=G^*). If a cursor satisfied this requirement we computed the optimal throttle sequence with that vector as a single pulse of length t_{opt} that accelerated the cursor to a maximum speed at half the distance between S and G^*, followed by a release of the throttle to allow the cursor's momentum to bring it to G^*, arriving at a velocity of 0. t_{opt} is estimated to the precision of our (60 Hz) screen resolution by finding the lowest number of frames (λ), such that: $\sum_0^\lambda D_\lambda^* > D/2$, where D is the Euclidean distance between S and G^* in screen coordinates and $D_\lambda^* = \sqrt{SS(\mathbb{P}_v * f(\lambda * 0.017))}$, where SS denotes sum of squares and $f(x)$ and $\mathbb{P}_v$ are from the above section describing task physics. Expressing optimal pulse length (t_{opt}) in frames (λ) automatically computes the number of units of fuel depleted by this optimal sequence. We subtract λ from 360 as our final estimate of the reward obtainable from the optimal route. (Note that we leave this score

on a scale of 0 to 360 for modeling purposes, but present score feedback to participants on each trial as a more intuitive proportion of preserved fuel).

If a cursor does not provide a single linear displacement solution, its optimal route instead comprises a two-pulse sequence using its two vectors that most closely align with the trajectory of the SG, i.e., the two vectors (v_1 and v_2) with the smallest "offset" values (θ_1 and θ_2) as computed in Figure 2b. The shortest combined displacement of these two vectors that moves a cursor from S to its most nearby Euclidean point on the circular threshold around G (G^{**}) can be computed by first originating v_1 at S and v_2 at G^{**} and finding where they intersect (∩). Forming an oblique triangle with lines |S∩|, |∩G^{**}| and |G^{**}S|, the length of |S∩| and |∩G^{**}| (i.e., the singular displacements of v_1 and v_2) can then be solved using the law of sines, i.e., $|S\cap| = |G^{**}S| * sin(\theta_2)/sin(60)$ and $|\cap G^{**}| = |G^{**}S| * sin(\theta_1)/sin(60)$. Optimal throttle sequence with these vectors is a vector of pulses (T_{opt}) containing [t_{v1}, t_{v2}], respectively solved with the lowest [λ1,λ2] values such that $\sum_0^{\lambda 1} D_\lambda^* > D_{v1}/2$ and $\sum_0^{\lambda 2} D_\lambda^* > D_{v2}/2$, where D_{v1} is the Euclidean distance between S and ∩, and D_{v2} is the Euclidean distance between ∩ and G^{**}. Given that λ2 is calculated from 0 velocity , the optimal sequence pulses v2 immediately upon the release of v1. We subtract λ_{total} from 360 as our final estimate of the reward obtainable from the optimal route, where $\lambda_{total}=\lambda 1+\lambda 2$.

In most cases λ_{total} is the same value whether using the above order, or by originating v_2 at S and v_1 at G^{**}, and estimating [t_{v2}, t_{v1}] relative to the resulting intersection (∩'). The exception occurs when one intersection (∩ or ∩') falls outside the grid, requiring more than one direction change to avoid catastrophic error with this sequence. However, all trials had at least one sequence with an intersection inside the grid for each cursor, i.e., at least one optimal path involving a single direction change. Our modeling framework simply required the lowest λ_{total} for each cursor on each trial, i.e., either λ from a single linear displacement, λ_{total} for either route if both intersections fall within the grid, or λ_{total} corresponding to the route with its intersection inside the grid, if one fell outside it.