Reaching vigor tracks learned prediction error

Colin C. Korbisch[1], Alaa A. Ahmed[1,2]

1. Department of Mechanical Engineering, University of Colorado Boulder
2. Biomedical Engineering Program, University of Colorado Boulder

Running head: Reaching vigor tracks learned prediction error

Address for Correspondence:
Alaa Ahmed
Neuromechanics Laboratory
Department of Mechanical Engineering
University of Colorado Boulder
Boulder, CO 80309-0354
email: alaa@colorado.edu

KEYWORDS: reaching; reward; value; decision making; movement vigor

GRANTS: Work is supported by grants from the National Institutes of Health (1R01NS096083) and the National Science Foundation (CAREER award 1352632) to AAA.

DISCLOSURES: The authors report no potential conflicts of interest.

AUTHOR CONTRIBUTIONS: CCK and AAA conceived and designed experiment. CCK performed experiments and analyzed data. CCK and AAA interpreted results of experiments. CCK prepared figures and drafted manuscript. CCK and AAA edited and revised manuscript. CCK and AAA approved final manuscript.

ORCiD Numbers: CK:0000-0002-6566-2860; AA:0000-0002-1596-342X

Institutional ID: http://ror.org/02ttsq026

ABSTRACT

Movement vigor across multiple modalities increases with reward, suggesting that the neural circuits that represent value influence the control of movement. Dopaminergic neuron (DAN) activity in the basal ganglia has been suggested as the potential mediator of this response. If DAN activity is the bridge between value and vigor, then vigor should track canonical mediators of this activity, namely reward expectation and reward prediction error. Here we ask if a similar time-locked response is present in vigor of reaching movements. We explore this link by leveraging the known phasic dopaminergic response to stochastic rewards, where activity is modulated by both reward expectation at cue and the prediction error at feedback. We used probabilistic rewards to create a reaching task rich in reward expectation, reward prediction error, and learning. In one experiment, target reward probabilities were explicitly stated, and in the other, were left unknown and to be learned by the participants. We included two stochastic rewards (probabilities 33% and 66%) and two deterministic ones (probabilities 100% and 0%). Outgoing peak velocity in both experiments increased with increasing reward expectation. Furthermore, we observed a short-latency response in the vigor of the ongoing movement, that tracked reward prediction error: either invigorating or enervating velocity consistent with the sign and magnitude of the error. Reaching kinematics also revealed the value-update process in a trial-to-trial fashion, similar to the effect of prediction error signals typical in dopamine-mediated striatal phasic activity. Lastly, reach vigor increased with reward history over trials, mirroring the motivational effects often linked to fluctuating dopamine levels. Taken together, our results demonstrate and exquisite link between known short-latency reward signals and the invigoration of both discrete and ongoing movements.

NEW & NOTEWORTHY

Previous research has demonstrated the invigorating effects of reward on movement. Growing evidence suggests this is causally explained by midbrain dopamine transients. Here, we demonstrate that reach vigor tracks canonical variables of learning and motivation across time scales ranging from milliseconds to minutes. Velocity was modulated by reward expectation, reward prediction error and reward rate, key variables that have also been associated with striatal dopaminergic fluctuations. These results point to a potential neural mechanism by which dopamine can influence both decision making and movement control and support the proposition that reward-based invigoration of movement is in part influenced by dopaminergic circuits.

INTRODUCTION

Imagine sitting down in front of a series of slot machines and being instructed to pull their levers, one after another. You find that these machines, frequently referred to as one-armed bandits for their ability to extricate money from patrons, have different rates of payouts, with some better than others. After learning these differences, you will likely prefer to pull the levers that offer a greater payout. But would you also choose to pull those levers faster or with more force?

Previous inquiry has shown that individuals will move faster towards goals or targets associated with greater value[1–8]. People are willing to produce greater muscle forces, or expend greater effort, to reach a cued target in less time. A potential explanation for how this may be represented in the brain lies in the neurotransmitter dopamine (DA), which is implicated in both the representation of value and the control of movement[9]. Basal ganglia output activity, already known to be influenced by dopaminergic inputs, has been found to invigorate not just movement vigor[10–12], but the decision-making process as well[13,14]. If vigor is indeed a reflection of value and DA the mediator, then vigor may also track the learning of value, due to the phasic dopamine release coincident with learning and reward prediction error[15]. Seminal work has shown that dopaminergic neuron (DAN) responses in a learned environment scale with reward prediction error at the time of both stimulus presentation and feedback presentation[16], but is also mediated by action-credit assignment[17]. DAN activity is greater in response to cues associated with greater expectation of reward (greater positive reward prediction error), and lower upon feedback of that reward (lower positive reward prediction error). If a similar time-locked response is present in movement vigor at cue and feedback presentation, this would suggest that DA-related activity contributes to reward-driven increases in vigor.

While short-term phasic DAN response is implicated in prediction error and learning, tonic dopamine levels have been found to be sensitive to the history of reward reception[1,18–20]. Particularly, midbrain DA levels come to match average reward rate and are implicated in motivation and invigoration. Hamid et. al. found a significant relationship between history of reward and relative DA in the nucleus accumbens compared to other potential analytes. Greater tonic levels of DA can also be interpreted as greater motivation or *drive*, resulting in behavioral invigoration[21,22]. Returning to our hypothetical gambler, receiving more rewards over the past, regardless of which levers were pulled, would result in greater vigor in subsequent lever pulls.

A growing body of work has found correlates between phasic dopaminergic activity and movement vigor[23–30]. In the 2024 study conducted by Engel et al., it was found that cue-evoked phasic dopamine response within rat nucleus accumbens core (NAc core) was significantly correlated with trial movement vigor, but only after the behavior had been sufficiently trained. Similarly, the 2018 da Silva work found that, for rat substantia nigra pars compacta (SNc) neurons sensitive to free movement initiation, relative activity levels were correlated with ensuing vigor. Optogenetic stimulation of these neurons in this study, however, did not produce any measurable changes in ongoing movement vigor, unlike the earlier Panigrahi study that found such an effect when striatal projection neurons were inhibited during a mouse lever task. These and other works provide evidence that not only do longer timescale tonic dopamine levels potentially modulate animal movement vigor, so too may short latency, phasic responses.

In this study, we sought to investigate whether human kinematic response to probabilistic rewards would mirror the characteristic reward prediction error response observed in mesencephalic

dopaminergic neurons (DAN) on a sub-second timescale. We also asked if kinematic response would reflect tonic DA response correlated with reward history. Given the short-latency and tuned response of DANs to reward value, we hypothesized vigor response in human arm reaching to be modulated by both the expectation of reward (probabilistic expectation) as well as the prediction error (difference between binary reward outcome and expectation). In effect, we expect outcomes congruent with two of the axioms for RPE-based models: 1) vigor responses to various reward expectations should demonstrate consistent lottery ordering, and 2) should show evidence of the "no surprise equivalence," i.e., fully deterministic reward outcomes show no relative difference . We also hypothesized that greater reward amounts received over the recent past would lead to greater relative invigoration, as a result of increased striatal DA levels. To test these hypotheses, we performed two experiments, each of which involved human subjects performing out-and-back reaching movements to probabilistically rewarding targets. Characteristic kinematics, including peak velocity and relative velocity of the return portion, were found to be better described by a learned utility estimate that integrated rewards and efforts, rather than by reward alone. These results highlight how motor planning, execution, and feedback control are all likely influenced by DA response, reflecting its dual roles in learning and motivation.

RESULTS

In a series of two experiments, we asked whether movement vigor tracked canonical determinants of dopamine-related learning and performance: learned value, reward prediction error, and reward history. In both experiments, human subjects directed arm reaching movements to targets associated with a probability of receiving a reward of a fixed value (1 point) or not (0 points) (Figure 1a-c; see Methods). Each of four target locations was associated with a unique probability of receiving the reward (0%, 33.$\bar{3}$%, 66.$\bar{6}$%, 100%) that was changed on each block (Figure 1d) of 180 trials. Feedback of reward reception or reward denial was provided upon target acquisition. Assuming perfect representation of task instruction, participants could experience one of five reward prediction errors (RPE): -0.33 and +0.66 for the 66% target, -0.66 and 0.33 for the 33% target, and 0 RPE for the 0% and 100% targets (Figure 1e). This paradigm allowed us to probe the effect of instructed vs. learned expectation of reward on the vigor of the outgoing movement. Additionally, we could investigate the effect of RPE at time of feedback on the vigor of the return movement. Lastly, we could test the effect of reward history on reach vigor across trials, independent of the expected value of the current target. We predicted that with increasing expectation, outgoing vigor would increase, reflected in an increase in peak velocity (Figure 1g). Additionally, we predicted that vigor would be modulated on the return portion of the reach, scaled by the reward prediction error, and vigor across trials would increase with reward history.

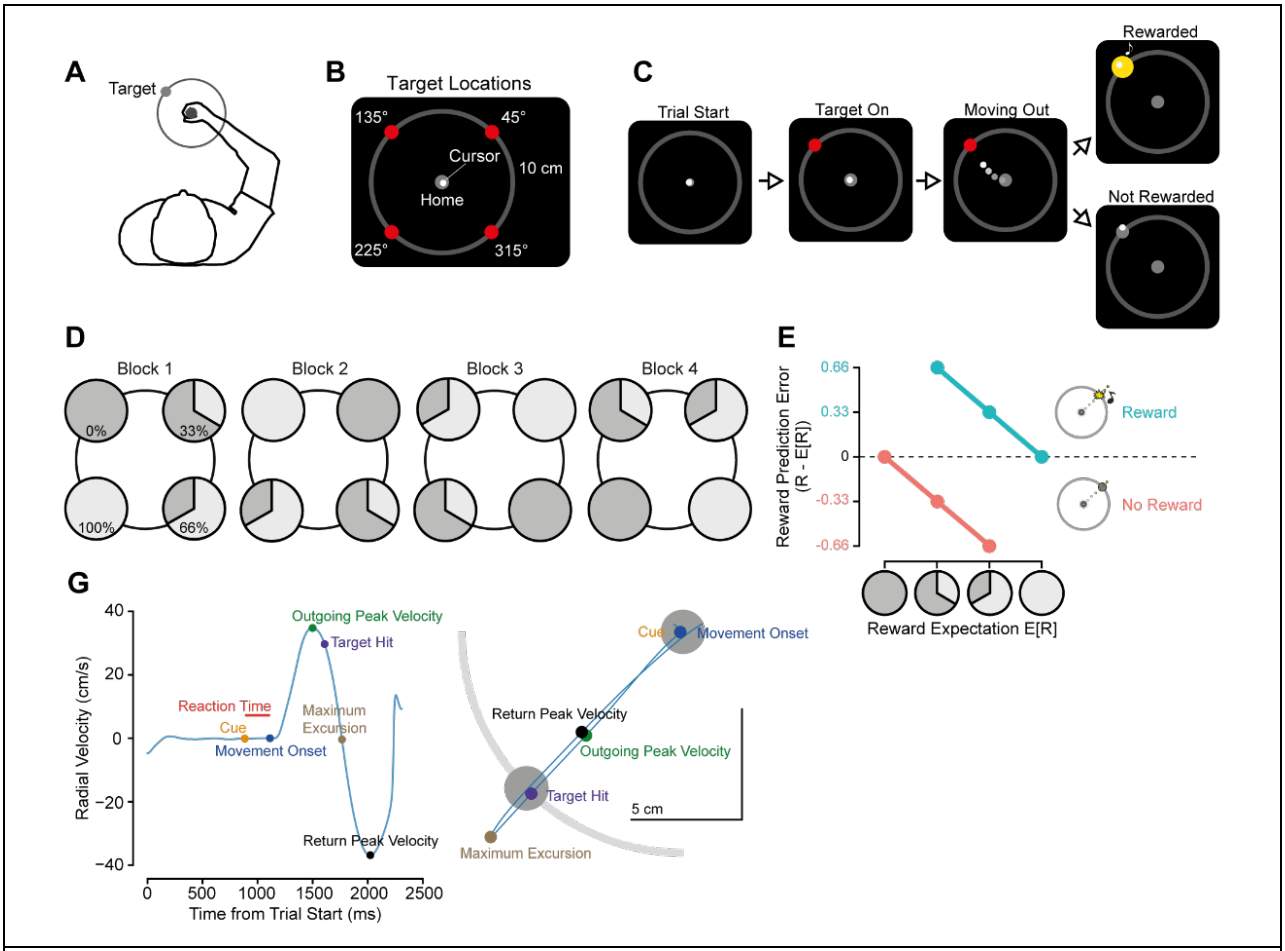

Figure 1. Experimental protocol. A) Participants made out-and-back reaching movements towards prompted, virtual targets. B) Targets were located at four different positions, each 10 cm from a central home location. C) Trial sequence. After the target was cued, participants were to make an out-and-back reach towards said target. Reward feedback was dependent on stated target reward frequency (in the first experiment), with reward feedback consisting of a yellow flash and high-pitched tone. Omission of feedback coincided with the target turning gray. D) At the beginning of each 180-trial block, in the first experiment, participants were told the reward frequencies of each target, either 0%, 33%, 66%, or 100%. Shown here is one possible experiment progression order. Relative order of reward-target associations was randomized across participants. Each block's rewards were pseudorandomized such that, in the single-target trials, the target reward frequencies were exact to the underlying frequency. Each block consisted of 5 sets of 36 trials, wherein each target was presented 9 times, and amount of reward per target was precisely equal to the stated expectation (i.e., rewards were given 0, 3, 6, or 9 times within each set of 36 trials depending on the target). Target presentation order within each set of 36 trials was fully randomized. E) Reward prediction error (RPE) was defined as the difference between the stated (and actual) reward frequency per target and the binary outcome of the trial. G) *Left* Kinematic measures of interest included the time between target presentation (cue) and movement onset (reaction time), peak velocity of the outgoing portion of the movement, time between movement onset and target hit, peak velocity of the return portion, and difference between the absolute value of the two peak velocities. Example positional trace of trial reach trajectory given on *Right*, with spatiotemporal points of interest highlighted.

Experiment 1

In the first experiment, we sought to focus solely on the vigor response to known expectation of reward. We did so by explicitly informing the participants verbally of the reward probability of each target at the beginning of the block.

Peak velocity to target tracks reward expectation

As the cued target's probability of reward increased, peak velocity of the outgoing movement to the target increased as well (GLMM, $\beta_{E[R]}$=0.0159±0.00629, p=0.0152, Figure 2a,b). Likewise, time to target, defined as the time between movement onset and reaching out to a 10 cm radial distance, decreased with increasing reward expectation ($\beta_{E[R]}$=-0.0125±0.00527, p=0.0226). It was also found that reaction times decreased with increasing reward expectations (Supplemental Figure 1a-b).

Return velocity is not solely predicted by outgoing velocity and instead influenced by reward prediction error

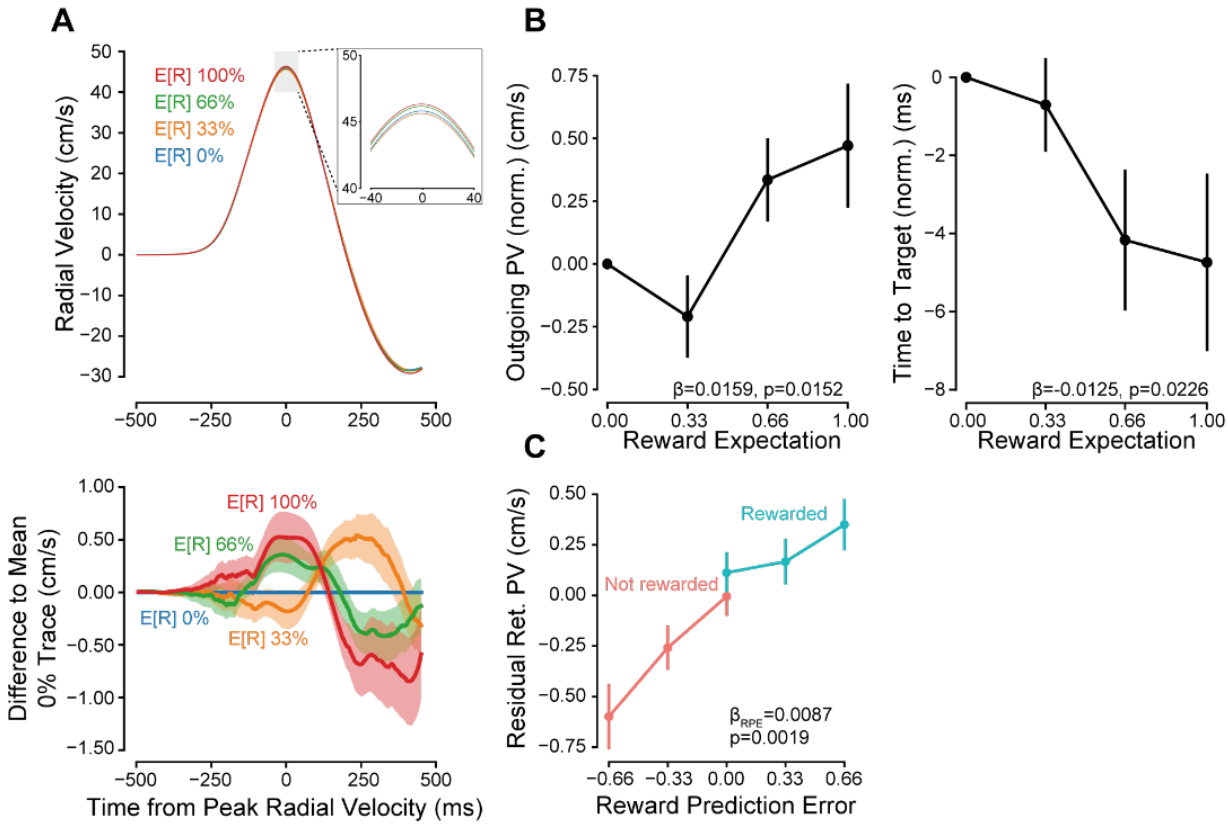

Figure 2. Vigor tracks reward expectation. A) *Upper* Radial velocity trace for first experiment, aligned to time of outgoing peak velocity. Highlighted region shows differences in peak excursion velocity for differences in reward expectations. *Lower* To calculate the difference traces, each participant's average signed radial velocity, at each sampled time point for the 0% expected reward condition, was subtracted from all other trials. Traces shown are the grand average of these difference traces, +/- standard error. B) *Left* Outgoing peak velocity increased with increasing expected reward. Data were normalized by subtracting the per-participant average for 0% expected reward condition. *Right* Similarly, the time to target hit, defined as the time between movement onset and reaching out to a 10cm radial distance, decreased as reward expectation increased. C) Return peak velocities were residualized by first fitting a reduced model including target direction and outgoing peak velocity. These residuals were significantly correlated with trial reward prediction error, increasing with greater RPE.

When the cursor hits the target, participants can either receive reward feedback or not. Does this feedback, or lack thereof, influence subsequent movement kinematics? Does the degree of surprise also modulate movement kinematics? In other words, does the reward prediction error influence the speed of the ensuing movement?

To answer these questions, we analyzed participants' return movements and whether receiving (or not receiving) reward feedback on a given trial produced significant influences. Specifically, we

questioned whether the sign and magnitude of the reward prediction error, $RPE,$ influenced the velocity:

$$RPE = R - E[R]$$

with R a binary coding for presence or absence of reward feedback (0,1), and $E[R]$ the instructed expectation of reward feedback (see Methods).

Participants exhibited diminished peak velocity on return movements compared to the outgoing movements (paired mean difference = 0.0601 m/s [0.0445,0.0757 95% CI], t_{41}=7.7789, p=1.34e-9). After controlling for the average relationship between outward and return peak velocities (β_{OutPV}=0.580±0.0192, p<2e-16), we asked whether there was an effect of reward feedback and reward expectation, i.e., RPE. In testing, we found a significant effect (β_{RPE}=0.00865±0.00260, p=0.00186), indicating that the return portion of the reach was not merely driven by feedforward mechanisms, but also feedback. No significant main effect of reward was found (β_{Reward}=0.000876±0.00159, p=0.586), nor interaction with RPE ($\beta_{RPE\times Reward}$=-0.00545±0.00392, p=0.171), signifying a slope of response across prediction error values without discontinuities. Additional control analyses confirmed return movements were modulated by reward feedback and not other potential confounds (Supplementary Figures 2-4).

To better account for variance in the outgoing portion of the movement, we normalized instantaneous velocity and focused on within-trial effects. First, instantaneous radial velocity was divided by the within-trial outgoing peak velocity. Next, the difference traces of these %-outgoing velocity values were calculated in a similar manner as in figure 2A, though instead of taking the difference to the 0% reward trials, the differences were taken relative to the per-participant normalized velocity trace for all 0% and 100% reward trials, i.e. RPE=0 trials. After calculating the grand average of these difference traces, we see a clear striation in relative normalized return velocity corresponding with reward prediction error (Figure 3a).

We next performed a hierarchical random effects analysis (Supplementary Figure 5; see Methods) to determine the population-average effect of reward prediction error on instantaneous, normalized velocity. At 212 ms after target feedback, we found a significant negative effect of RPE, indicative of greater invigoration in the return movement with greater RPE (Figure 3a).

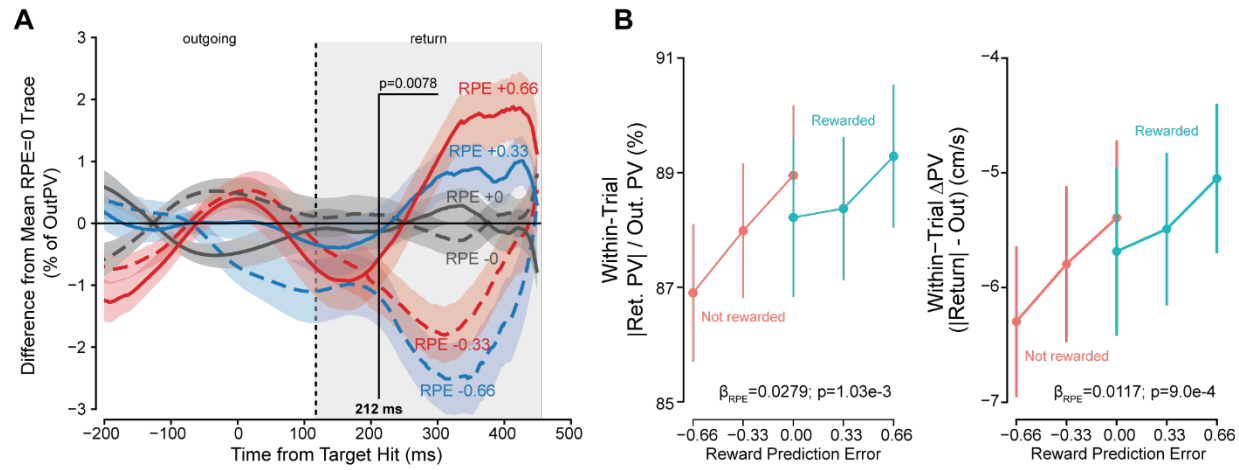

Figure 3. Vigor response reflects reward prediction error. A) For a within-trial measure of relative invigoration, instantaneous velocity was normalized by dividing by the trial outgoing peak. Difference traces were then calculated in a manner similar to figure 2A. From SPM analysis, a significant effect of RPE on percent peak velocity was found beginning at 212 ms after reward feedback was received. B) The within-trial difference in absolute peak velocities, *Left* return divided by outgoing *Right* return minus outgoing, significantly varied with reward prediction error. As RPE increased, the return peak more closely matched the outgoing peak.

Aside from the effect on comparative instantaneous velocity, the within-trial difference in velocities, i.e., the difference between excursion and return peak velocity, also exhibited a significant effect of RPE (Figure 3b); with the slope of response consistent across positive and negative RPE conditions ($\beta_{RPE \times Reward}$ =-0.00176±0.00502, p=0.727). For both percent relative to outgoing peak (3b *Left*) and velocity difference (3b *Right*) the slope of response was significant across RPE conditions (β_{RPE}=0.02788±0.00788, p=1.03e-3; β_{RPE}=0.0117±0.00325, p=9.00e-4, respectively). When RPE was most positive (+0.66) the difference between excursion and return peak velocity was at its smallest compared to RPE at its most negative (-0.66). There was no significant main effect of reward feedback in either metric (β_{Reward}=-0.00269±0.00463, p=0.5643; β_{Reward}=-0.00355±0.00208, p=0.0978).

Overall, we found a significant effect of reward feedback on the return movement that was dependent on the reward expectation, i.e., the reward prediction error. With greater RPE, relative return velocity was greater compared to when the prediction error was negative. This difference emerged 212 ms after feedback was received and could not be accounted for by variation in maximum excursion, nor preemptive differences in the outgoing portion of the movement (potential correlations between outgoing peak velocity and future reward reception). Participants employed neither set return velocity, nor set return velocity difference strategies. In the first case, outgoing peak velocity would be unrelated to the return peak, and in the second case, return velocity difference would be invariant to the experienced prediction error.

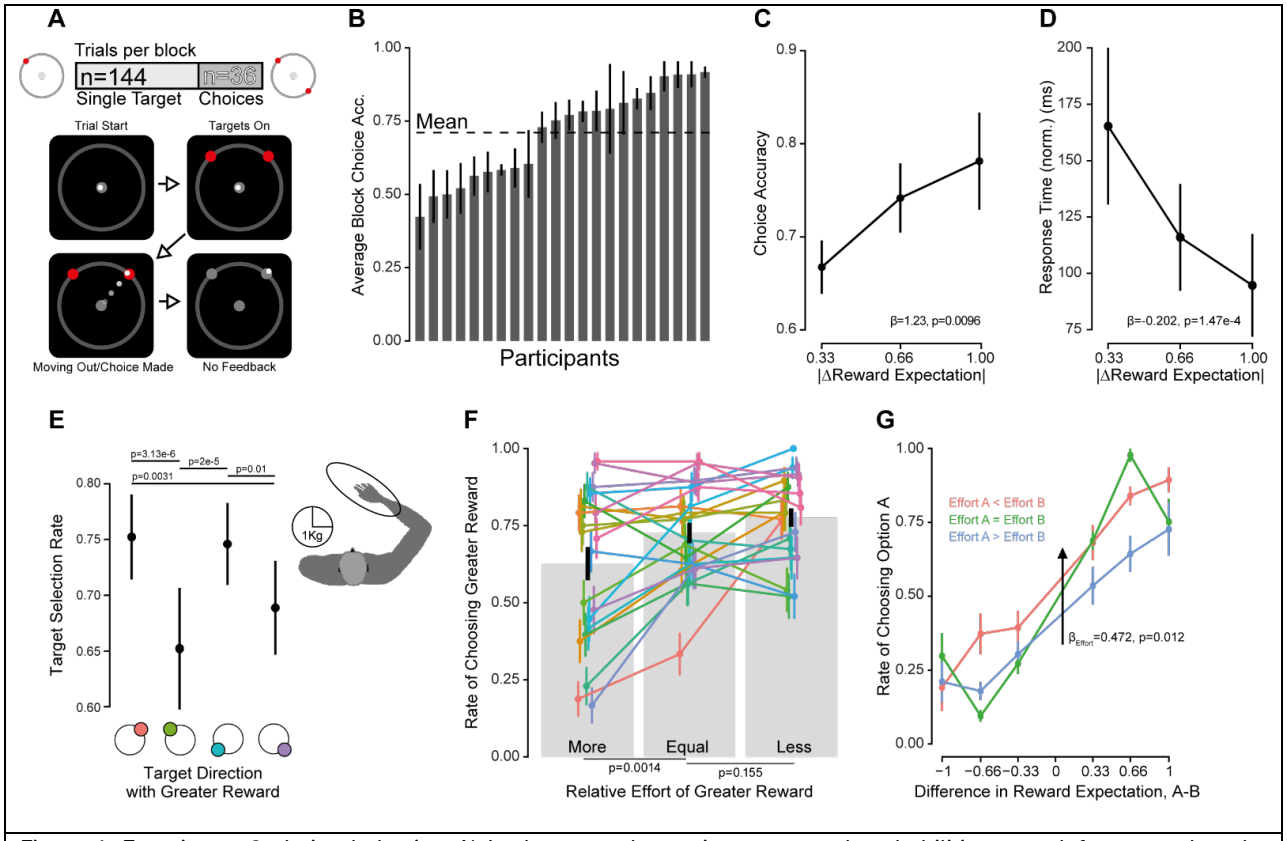

Figure 4. Experiment 2 choice behavior. A) In the second experiment, reward probabilities were left unstated to the participants. Instead, they were incentivized to learn the underlying rewards during the single-target period, as additional monetary compensation was dependent on performance during later choice trials (see Methods). B) Per-participant average accuracy rate over four blocks. Participant choices were highly variable across the population. Mean accuracy was 71%. C) Participant choices were found to be sensitive to relative expectation, choosing the more rewarding option as difference increased. D) Response time decreased with increasing difference of reward expectations of the presented targets. E) Target direction was effected choice selection. When targets 1 or 3 were the more rewarding option, i.e. associated with lesser biomechanical effort, they were selected at higher rates compared to when targets 2 or 4 were the more rewarding option, consistent with a potential influence of biomechanical effort on choice selection. F) When grouping choices by whether the greater reward has greater relative effective mass, significant differences in choice accuracies were found. Individual participant responses are shown in color. G) Baseline selection rates were shifted depending on relative effort. No interaction between effort and reward difference was found.

Experiment 2

To further investigate the effects of probabilistic reward and test the degree to which individuals responded to implicit, learned value, we conducted a second experiment where participants were left uninformed of the targets' reward frequencies. Instead, participants learned through experience over the course of a block, affording us a measure of trial-to-trial response of changes in reward estimation. Within each block, after a sequence of single-target trials similar to the first experiment, there were 36 two-alternative forced choice trials (2-AFC) to assess degree of learning (Figure 4a). Individuals were incentivized to choose greater expected rewards on choice trials as additional monetary compensation was dependent on performance (see Methods).

Choices reflected both expectation of reward and effort

The choice trials at the end of each block allowed us to determine how well the participants had learned the reward expectation of each target. Critically, no feedback was given on target hit during these choice trials to limit learning to the single target trials only. Choice accuracy, defined as the

selection of the greater reward frequency of the two options, varied across the participants, ranging from 42.7% to 91.7% (Figure 4b). The population-average accuracy rate was 71%. Underlying reward difference, hidden to participants, was found to significantly predict rate of choosing one option over another ($\beta_{\Delta E[R]}$=1.233±0.476, p=0.0096; Figure 4c). Response time of the decision was found to significantly decrease with increasing difference of reward expectation (Figure 4d; $\beta_{|\Delta E[R]|}$=-0.202±0.053, p=1.47e-4), providing further evidence that underlying expectations were learned. Interestingly, response time and outgoing peak velocity response differed during choice trials in their sensitivity to reward difference and choice accuracy (Supplementary Figure 6). Whereas outgoing peak velocity only varied with the selected option's reward expectation, response time varied with both relative and selected expectation.

Target direction was also found to influence decision making. We first categorized all choices by whether a given target direction was associated with the greater of the two potential rewards presented on a given choice trial. We then calculated the choice accuracy rate for each of these subsets. If direction had no influence, we should expect no variation in selection rates between these four categories, however this was not the case. Instead, significant differences were found (Figure 4e). Selection rate biases matched the arm's inertial axes[31] (Figure 4e *Right*). Using this measure of effort, we categorized choice trials by whether the more rewarding option was associated with more or less effort compared to the alternative choice (Figure 4f). We found significant differences in accuracy rates, with idiosyncratic responses to effort readily apparent. Augmenting our logistic regression analysis from before by including a relative effort term, we found a significant effect of said relative effort, with average rates of choosing the more rewarding option increasing when this option was associated with less effort (Figure 4g). Thus, participants' choices revealed they had largely learned the target reward contingencies and demonstrated that these value-based choices were influenced by the effort of the arm reach.

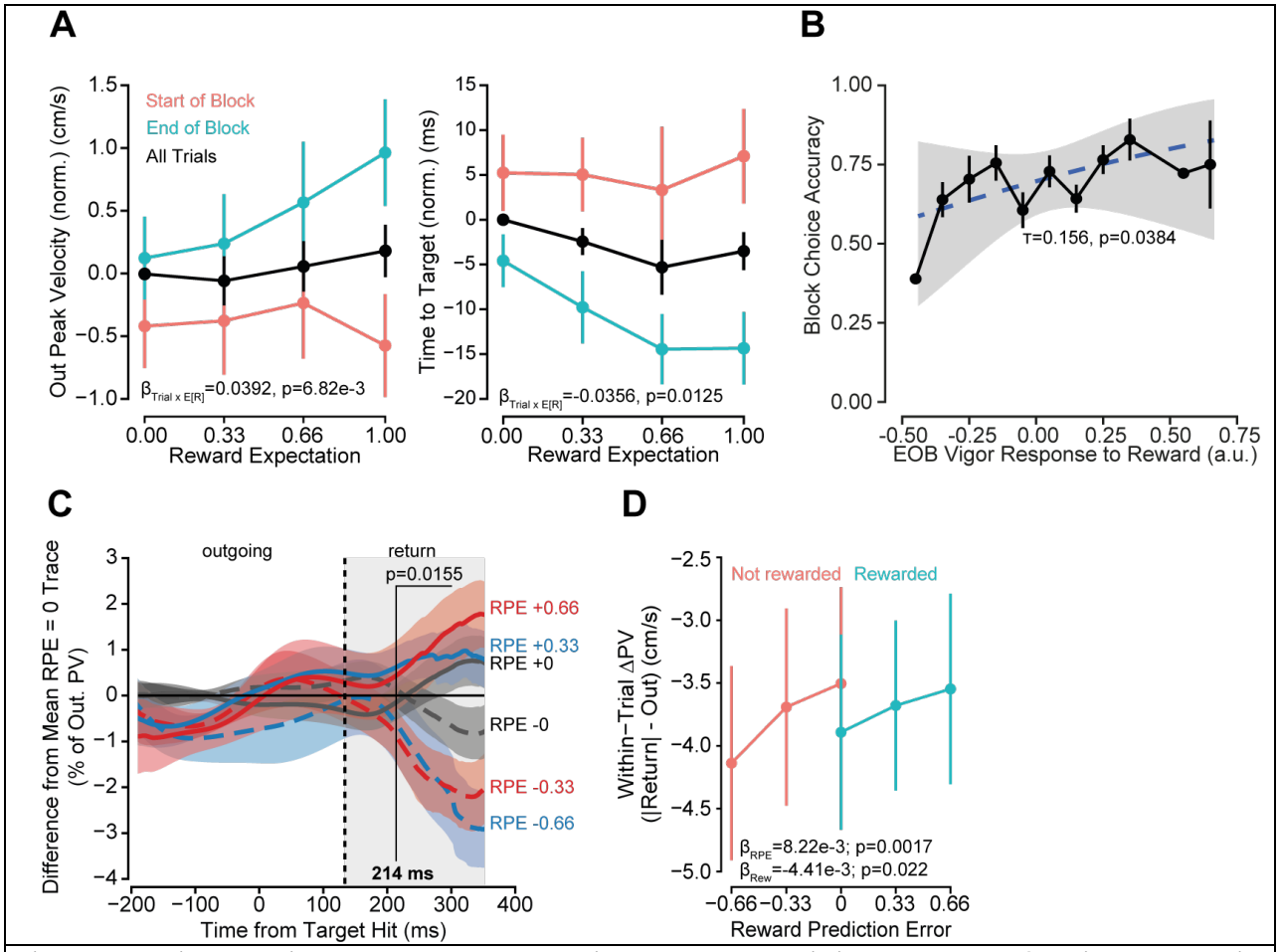

Figure 5. Experiment 2: Vigor tracked reward expectation and reward prediction error. A) *Left* Outgoing peak velocity increased with increasing reward expectation. Data were normalized by taking the difference to per-participant average for 0% reward. Points and error bars represent grand means, ± standard error. The slope of peak velocity response to reward expectation changed over the course of a block. In the initial 36 trials (start of block), response was essentially flat, but by the final 36 trials (end of block), outgoing velocity significantly increased with respect to reward frequency. *Right* Time to target, taken as the duration between movement onset and reaching out to a 10 cm radius (target hit) changed with respect to underlying expectation over the course of the single-target period. B) End of block response to reward (i.e. per-participant slope of response to E[R]) was significantly correlated with block choice accuracy (rate of choosing the more rewarding option, Figure 4B). C) Normalized velocity difference traces for experiment 2. As in figure 3C, velocity data were first normalized by dividing by the within-trial peak outgoing velocity. Difference traces were calculated by subtracting from each trial the average instantaneous normalized velocity for trials where the underlying reward expectation was 0% or 100% (RPE = 0). Data presented is the grand average of these difference traces. Differences in relative velocity emerged after reward feedback that scaled with prediction error. D) As RPE increased, the relative reduction in peak velocities between the return and outgoing portions of the reaching movement decreased (i.e. the difference was less negative).

Velocity in single-target trials tracked reward expectation and reflected learning

Given that participants' choices revealed learned relative reward value, we asked whether vigor on single target trials would reveal this value estimation as learning progressed. Over the course of the single-target period, we found that average slope of response of outgoing peak velocity relative to reward expectation increased (GLMM; $\beta_{Trial \times E[R]}$ = 0.0392±0.01447, p=0.00682), demonstrating a learned response to the probabilistic reward (Figure 5a). At the beginning of a block, there was no significant effect of reward expectation ($\beta_{E[R]}$= -0.0108±0.0073, p=0.142). Likewise, the effect of expectation on the time from onset to reaching the 10 cm radius (time to target) significantly varied over this period ($\beta_{Trial \times E[R]}$ = -0.0356±0.0142, p=0.0125). The change in response of instantaneous

velocity over the course of the single-target period is further evident in the velocity difference traces (Supplementary Figure 7). However, no such interaction effect between trial and hidden reward expectation was found when modeling reaction times (Supplementary Figure 1).

The strength of peak velocity-reward expectation response (i.e. the slope of velocity relative to reward) could predict within-block choice accuracy (the rate of choosing the more rewarding of the two options). Across the participant pool, we found a significant correlation between these measures (Kendall's rank correlation, $\tau = 0.156$, p=0.0384; Figure 5b). The stronger the slope of response in an individual at the end of the single target period, the greater the rate in selecting the more rewarding of the two options presented during choice trials.

Even when participants were not told of the expected reward frequencies at the outset of a block, both outgoing peak velocity and time to target responded to increasing expectation. Learning of the different reward frequencies is shown clearly by the differential response of velocity at the beginning and end of the single-target period. Initially, kinematic response did not differ between the four targets, but by the end, average peak velocity increased with increasing expectation of reward feedback. Within-block performance could be predicted based on participants' slope of response to reward at the end of the single-target period, suggesting that the change in kinematics over the course of a block was related to the learning of relative reward frequencies.

Change in return velocity tracks reward prediction error

Turning to within-trial measures of relative vigor, we performed the same hierarchical random effect SPM analysis on normalized velocity as in the first experiment and found a population-level effect of RPE (Figure 5c). Additional supporting detail for this analysis is provided in supplementary documentation (Supplementary Figure 8). At 214 ms after reward feedback, relative velocity varied significantly with reward prediction error. Likewise, the within-trial velocity difference (return minus outgoing) also significantly varied with RPE (β_{RPE}=8.22e-3±2.61e-3, p=0.00166; Figure 5d). With increasing positive prediction error, the return peak velocity more closely matched the outgoing peak velocity. Reward feedback had no significant interaction effect on this difference ($\beta_{RPE \times Reward}$=-2.156e-3±3.682e-3, p=0.558), However, there was a significant main effect of reward (β_{Reward} =-4.414e-3±1.825e-3, p=0.0217). As such, for both 0% and 100% rewarding trials, the differences in outgoing and return peak velocities were statistically distinct. The slope of response to reward prediction error was not found to change over the course of a block ($\beta_{RPE \times Trial}$=-4.746e-3±4.770e-3, p=0.320). As in the previous experiment, we performed additional control analysis to isolate the effect of reward feedback on the return movement (Supplementary Figure 9).

In summary, even when expected reward must be experienced and learned, individuals change their behavior on the return movement in a manner consistent with RPE sign and magnitude.

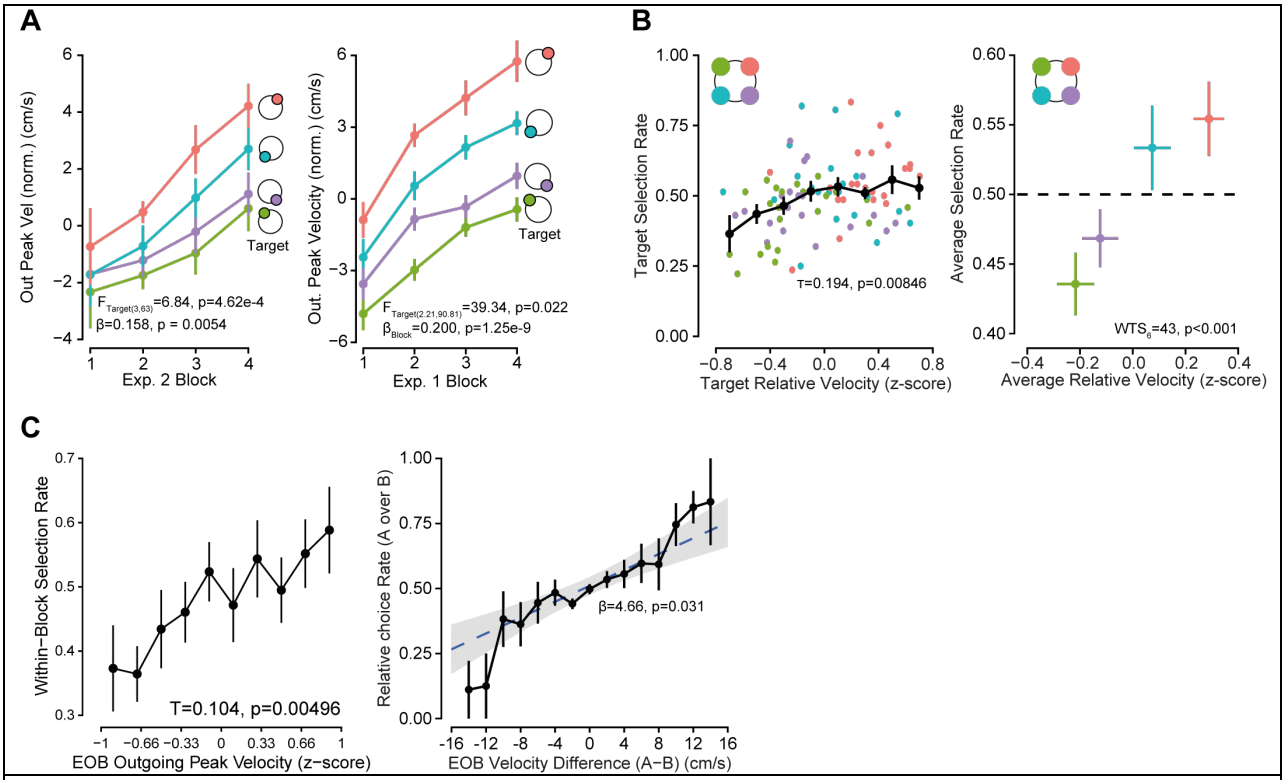

Figure 6. Relation between velocity, subjective effort, and choice preference. A) *Left* In experiment 2, arm reaches towards target 1 were found to have a greater peak velocity on average, with targets 2 and 4 associated with comparatively lesser averages. *Right* In the first experiment, the same ordering of relative peak velocities was observed. B) *Left* Participant-relative velocity, in single-target trials, were correlated with relative rates of selection of the different target directions. *Right* Averaged over all participants per target, there was evidence of a separation in relative average velocities and selection rates. C) *Left* Relative velocity to a given target at the end of the single-target period could predict preference. *Right* Difference in average velocity between two targets, within the final 36 single-target trials, was continuous with relative rates of selection for those two targets. The greater the difference in velocity, the more frequently one option was chosen over the other in the proceeding six choice trials where the two targets were presented together.

Biomechanical effort slowed outgoing peak velocity

Considering the effect of target direction (i.e. relative effort), on choice bias, we were also concerned whether there would be differences in average peak velocities towards the different directions. Investigating this potential effect (Figure 6a *Left*), we found that each direction was significantly different in average peak outgoing velocity from the other (Tukey-HSD comparison testing; Supplementary Table 2). In addition to the directional effect, we found that average outgoing peak velocity increased over the course of the experiment from block-to-block (β_{Block}=0.158±0.051, p=0.0054).

We also evaluated whether this directional influence on velocity was conserved in the previous experiment (Figure 6a *Right*). Once again, differences in average outgoing velocity across the directions were consistent with the effective mass of the arm. Peak velocity increased over the course of the experiment (β_{Block}=0.2352±0.0058, p=1.25e-9) for all targets, and post-hoc multiple comparison testing (Tukey HSD) revealed significant differences across all four targets (Supplementary Table 1).

In short, average velocity to the different target directions in both experiments reflected the relative inertia, or effective mass, of the arm when making this reaching movement. Outgoing peak velocity was greatest towards the 45° target, and slowest towards the 135° target.

Subjective vigor response in single-target trials predicted choice

Summarizing our results thus far for the second experiment, we have shown how participant choices reflected subjective value, and kinematic response during single-target trials reflected the learning of relative reward frequencies. To further substantiate the relationship between vigor and preference, we examined the degree to which reach velocity during the single-target trial period could predict choices across our participant pool.

To start, we analyzed whether the subjective effort bias evidenced in participant choices would also be present in the kinematic response. Comparing the per-participant relative velocities and the target selection rates, we found a significant correlation (Kendall's rank correlation, τ = 0.194, p=0.00846; Figure 6b *Left*). With MANOVA analysis, after averaging across participants, we found significant differences between the relative velocities and selection rates across the four target directions (Figure 6b *Right;* $WTS_{(6)}$=43.165, p<0.001). Next, we questioned whether kinematic behavior during the end of the single target period (the final 36 single-target trials) could predict later choice selections directly. To do so, we analyzed the correlation between the relative peak velocity per-target, per-participant and that target's within-block selection rate (quantified as a fraction of the total 18 potential selections within the choice period). A significant correlation was found (Kendall's rank correlation, τ = 0.104, p=0.00496; Figure 6c *Left*). The faster someone reached towards a target at the end of the single-target period, the more frequently they chose that option during choice trials.

As an additional piece of evidence for the relationship between outgoing peak velocity and relative selection rate, we directly compared the difference in mean velocity between two targets at the end of the single target period and the relative rate of selection within that block. Via logistic regression analysis, we found a significant relationship between the velocity difference and rate of selection ($\beta_{\Delta PV}$ = 9.56±2.55, p=1.75e-4; Figure 6c *Right*). Roughly, a 0.2 m/s difference in peak reach velocity translates to a 3:1 preferred rate of selection.

Taken together, we found that participant decisions varied widely across the population but could be predicted by each participant's relative outgoing velocities on single target trials: the greater the relative peak velocity, the more frequently that option was chosen. This trend held both for the relative reward values as well as the target directions.

<u>*Value Estimation*</u>

Reward and effort (in the form of target direction) had significant influences on excursion velocity. Both reward and effort also affected choice preference and relative selection rates. We posit that the target characteristics of reward expectation and direction could be combined into a singular subjective value or decision variable that would describe both participant-specific reach velocity and choice preference.

Learned value better explained single-target trial reach vigor

Final reach behavior at the end of learning (at the end of the single trial block), could explain participant choices in the following choice trials. If vigor indeed reflects subjective value, then we should see this learned value emerge on a trial-to-trial basis in the reach behavior during the single target trials. To examine whether this was the case, we fit a Bayesian hierarchical delta-rule learning model, similar to a Rescorla-Wagner model [32], to choices at the end of each block to determine the

per-trial subjective value for each participant as they learned these values over the course of the single target trials. After each trial, the updated value estimate of a target V'^T is calculated:

$$V'^T_i = V^T_i + \eta_s\left(R_i + e^T_s - V^T_i\right)$$

Here, R_i is the reward received on the current trial, η_s is learning rate, and e^T_s, the relative directional effort of the target. Participant-specific parameters, η_s and e^T_s, were selected and sampled from a population distribution in a hierarchical fashion. A single relative directional effort parameter was fit for each subject and each target direction.

A separate logistic regression confirmed that the model-derived value differences could predict population-level aggregate choice behavior (Figure 7a). As relative value increased, individuals were more likely to choose that option. Additionally, the model-derived learned value, which incorporated both individual effort valuation and trial-to-trial reward value updates, was found to better predict choices as compared to the objective, underlying reward expectation, with an aggregate accuracy rate of 75.4% compared to 70.98%. This was corroborated with a significant difference in ROC curves between the two metrics (Bootstrap test for difference in AUC: D=8.054, n=2000, p=8.014e-16). However, choice prediction improvements, relative to simple reward expectation difference, were highly idiosyncratic (Figure 7b). Choice trial response time was significantly influenced by value, decreasing with increasing value difference ($\beta_{|\Delta V|}$=-0.1373±0.0464, p=0.0094; Figure 7c).

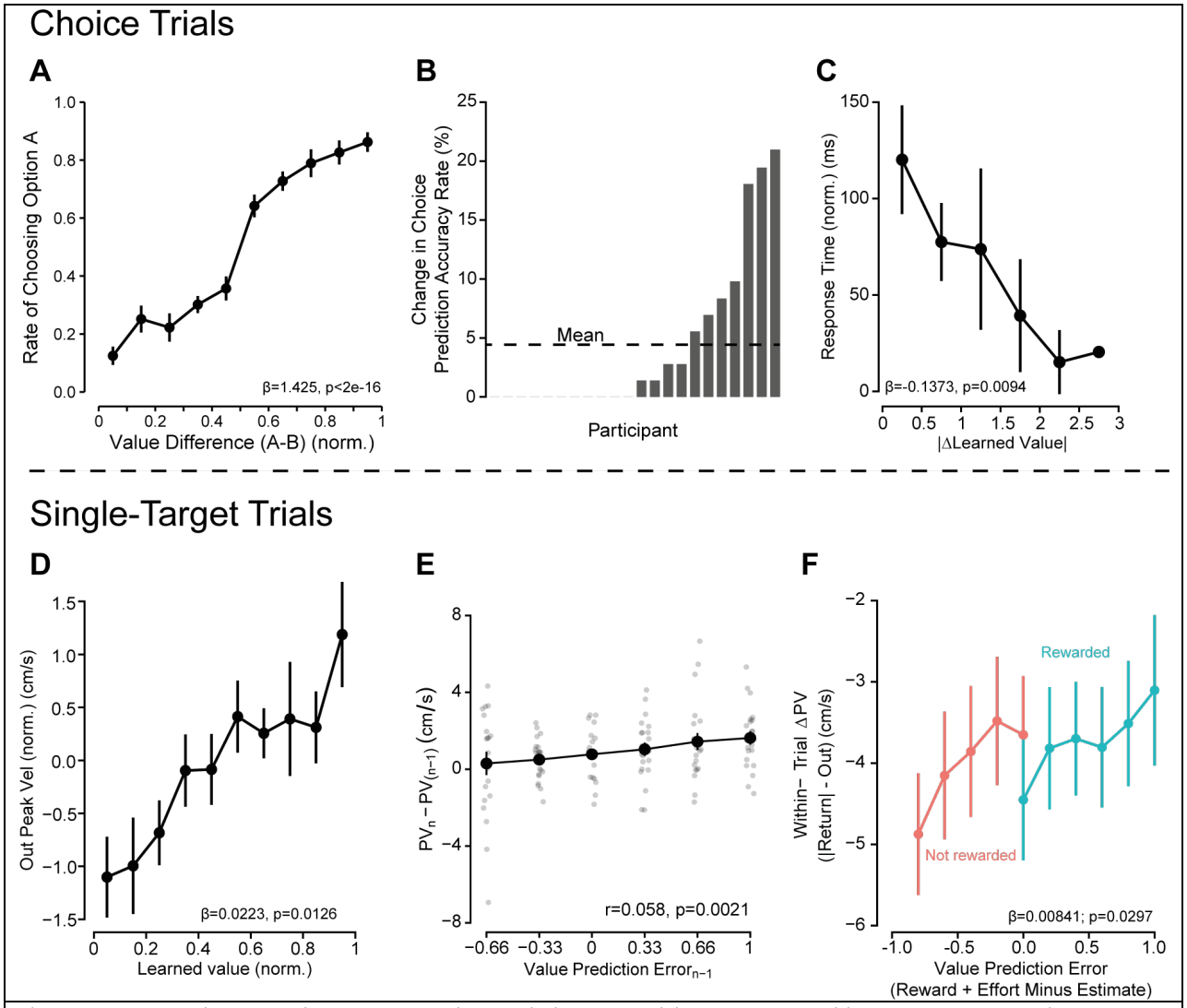

Figure 7. Model-derived value improved behavior predictions. A) Individuals were sensitive to learned value differences during decisions, selecting targets with greater relative values. B) Choice prediction accuracy improvement was participant specific. Average improvement was 4.42%, and some participants saw choice prediction improvements up to 20% (roughly 29 choices during the experiment). C) Choice trial response time significantly decreased with difference of the presented targets' learned values. D) Normalizing value relative to per-participant minimum and maximum estimate (min-max norm.), we found that outgoing peak velocity during the single-target trials significantly increased with increasing learned value. E) On trials where the same target was prompted in repetition, the previous trial's reward prediction error was significantly correlated with the trial-to-trial change in outgoing peak velocity. F) Value prediction error, taken as the difference between feedback and expected value, was also found to predict the relative difference in peak velocity between outgoing and return.

We next used this participant-specific trial-to-trial learned value to predict reach peak velocity on each single target trial. We found a significant relationship between estimated value of a target and outgoing peak velocity (GLMM; β_{Value}=0.0223±0.00895, p=0.0126; Figure 7d). The regression model including value improved in performance relative to one using E[R] (AICc: -13479.4 versus -13369.3 respectively). As subjective value increased, so did peak velocity of the reach. Time to target decreased with learned value as well, and better matched the data compared to reward expectation (β_{Value}=-0.0265±0.007148, p=0.0013; Figure S6B; AICc: -12225.6 versus -12157.7).

Interestingly, even though subjective directional efforts were incorporated into the target value, there were still significant effects of direction. Post-hoc Tukey HSD testing revealed significant differences in average outgoing peak velocity between targets 1 and targets 2 and 4 (Supplementary Table 3). This

implies that the difference in vigor between the target directions does not fully capture the resultant difference in value as shown by later choices.

Additionally, if our model-estimate for value were related to the outgoing peak velocity, we predicted that an individual's change in value (the value prediction error) would be correlated with the trial-to-trial change in peak velocity when target direction is repeated (Figure 7e). Calculating this correlation, we found that, as hypothesized, the trial-to-trial difference in outgoing peak velocity, when direction is repeated, is significantly correlated with the previous trial's value prediction error (Pearson's correlation, $t_{(2801)}$=3.083, r=0.058, p=2.07e-3). Testing to make sure this effect was target-specific, we calculated this correlation on trials where relative effort, but not direction was repeated, and found no significance ($t_{(2732)}$=-0.624, r=-0.012, p=0.533). Thus, the change in learned value for a specific target was reflected in the change in reach velocity from one trial to the next.

The relative difference in peak velocities between the outgoing and return portions of the movement decreased with increasing value prediction error (β_{VPE}=0.008414±0.00370, p=0.0297; Figure 7f); the comparative velocity of the return portion of the reach was greater with more positive error. Performance in modeling the within-trial velocity difference also improved when using VPE compared to RPE (AICc: -37118 versus -37098.2). Unlike the RPE regression model that showed a significant effect of reward reception on average difference in return velocity (see above, Figure 5d), the value prediction error model did not (β_{Reward}=-0.00441±0.00296, p=0.153). Additionally, no significant interaction was found ($\beta_{VPE\ x\ Reward}$ = -0.00510±0.00298, p=0.0873). In effect, the slope of response of velocity difference to value prediction error remained constant with or without reward feedback. Taken together, relative return velocities varied continuously with value prediction error with no additional significant effect of reward reception.

We were also interested in whether the participant-specific estimate of learned value would better inform our analysis of kinematic behavior in the choice trials (Supplementary Figure 10). We found again that outgoing velocity significantly varied with the learned value but was not influenced by the alternative option's value (β_{V_Hit} =0.0339±0.01448, p=0.030; β_{V_Rej}=0.00269±0.00605, p=0.656). As with previous kinematic measures in single-target trials, regression model performance improved relative to the use of latent reward expectation (AICc: -2645.3 versus -2611.1). The time to target decreased with the chosen, learned value as well (β_{V_Hit} =-0.05784±0.01477, p=0.00111). As stated previously, response time decreased as value difference increased; however, this measure was an exception to regression model performance improvements and information criteria. Value difference showed a marginal decrease in AICc value compared to reward difference, 3289.8 versus 3268.2. However, looking further, conditional and marginal R^2 were improved with value difference: 0.334/0.015 versus 0.332/0.009. In effect, value difference better matched the experimental data, but lacked predictive power compared to reward difference for hypothetical population data.

Recent history of reward led to faster movements

Having shown how target-specific learned value influenced the reach kinematics on the outgoing and return portions of the movement, we sought to investigate the potential effects of target-agnostic reward reception, independent of value estimation. Previous research has found that midbrain dopamine levels fluctuate with differing levels of average reward received in recent history[18], justifying our analysis. We investigated the potential effect of target-independent reward history on participant kinematics in both experiments, testing any potential influence of immediate previous

trial reward reception (regardless of expectation or effort), as well as an integrated reward history making use of a "leaky integrator" model[33] (see Methods). Based on previous findings, we hypothesized that reward history would invigorate the outgoing movement, increasing peak velocity, and decrease response times. We modeled the reward history term using the following equation:

$$\bar{R}_{i+1} = \alpha R_i + (1-\alpha)\bar{R}_i$$

Where the updated reward history value $\bar{R}$ was dependent on the received reward feedback R and the smoothing factor α.

In our first experiment, after controlling for the immediate trial's influence (reward expectation and directional effort), we also found an effect of previous reward reception and peak velocity. On trials

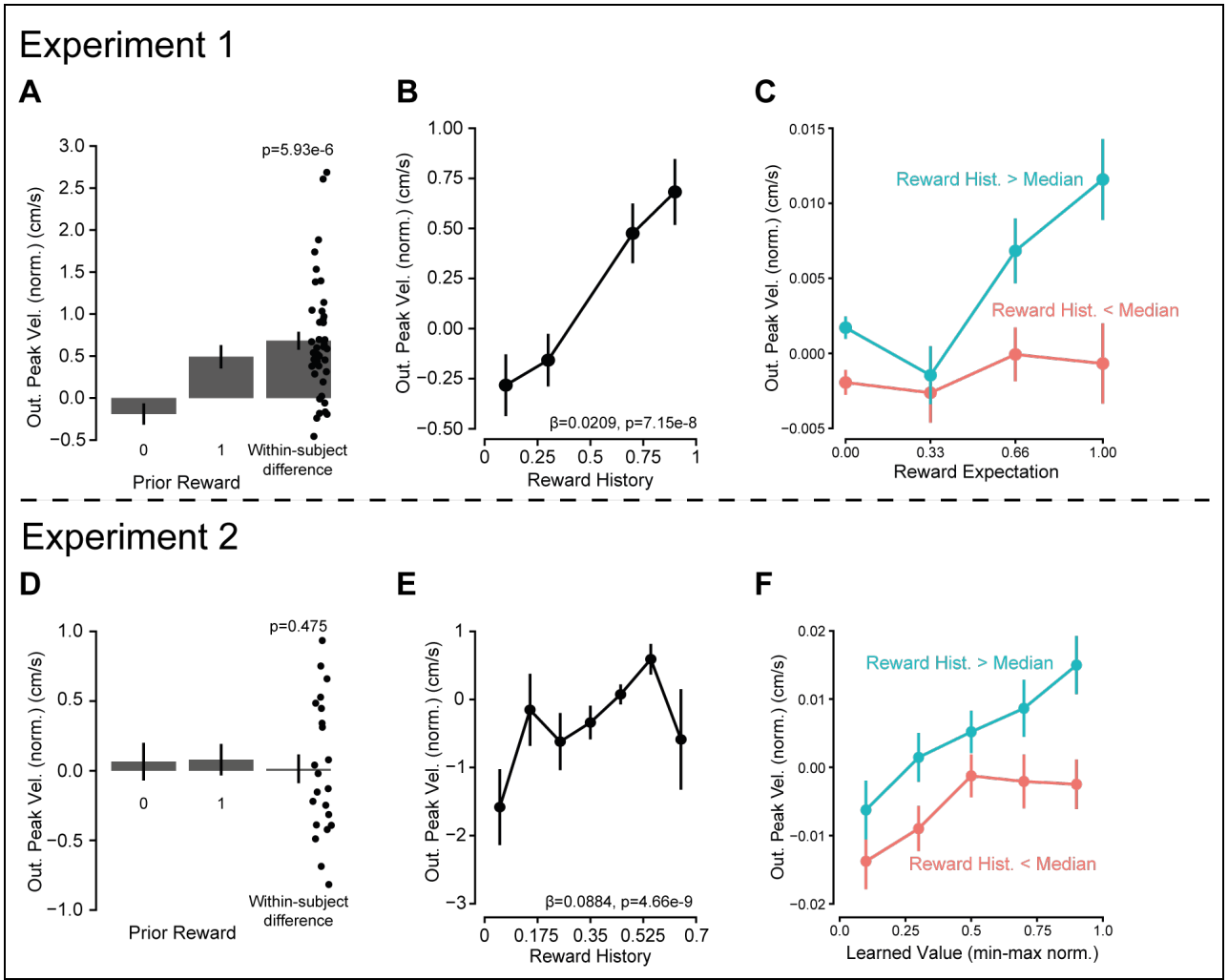

Figure 8. Peak velocity increased with reward history. Experiment 1: A) Outgoing velocity significantly increased with prior reward reception B) Reward history was estimated based on an exponentially smoothed moving average fit to reaction time (see Methods). With greater reward history, outgoing peak velocity increased. C) Even when accounting for movements to the same reward expectation, reward history has a significant effect on outgoing peak velocity. Experiment 2: D) Prior reward reception did not influence the subsequent trial velocity. E) The effect of reward history was determined in the same manner as in experiment 1. Outgoing peak velocity significantly increased with the amount of reward received in recent trial history. F) Vigor of reaching movements towards targets of similar learned value, were influenced by recent trial reward history.

following reward, regardless of what the probability of reception was, peak velocity was significantly greater compared to trials following no reward (β_{PriorR}=0.0152±0.003, p=5.93e-6; Figure 8a). There was no significant interaction between this prior reward effect and current-trial reward expectation (p =0.713).

To further analyze reward history and the integration of previous reward reception within a block, an exponential smoothing model was applied to quantify local average reward expectation. After fitting to reaction times, the smoothing factor, α, was found to equal 0.647. As expected, reaction time significantly varied with history, though contrary to our hypothesis, increasing rather than decreasing ($\beta_{History}$=0.01295±0.00248, p=1.77e-7). Incorporating reward history into our previous regression model for peak velocity, we found that excursion peak velocity significantly increased with increasing $\bar{R}$ ($\beta_{History}$=0.0209±0.00319, p=7.15e-8; Figure 8b). Model performance also improved with respect to information criteria (AICc: -85753.7 versus -85728.6). We found that this model-derived reward history value was significant even when accounting for current trial reward expectation (Figure 8c). An exemplary participant and block is shown in supplementary figure 11, depicting how reward history values are updated with rewarding experience. Average trial counts per participant for each binned reward history value is shown as well, showing that distributions were roughly uniform.

To summarize, in our first experiment, we found that peak velocity was influenced by the recent trial history. Reward reception on the immediate prior trial invigorated movement, leading to greater velocities. In fitting a "leaky integrator" model to reaction time, we found that recent reward context, beyond the immediate prior trial, led to greater peak velocity.

In our second experiment, unlike the first, there was no significant response caused by reward received on a previous trial (β_{PriorR}=-0.0033±0.006, p=0.592; Figure 8d). To see if there was any effect of history at all, we again modeled whether an exponential moving average model could predict an effect of reward history on a given trial's behavior. With the same reaction time dependent likelihood maximization, we found a smoothing factor, α, equal to 0.0521. After fitting, inclusion of the reward history value into the full reaction time regression found a significant effect ($\beta_{History}$=-0.160±0.0246, p=8.78e-11): reaction time decreases with history of reward. Outgoing peak velocity again significantly increased with reward history (β_{Hist}=0.0884±0.0151, p=4.66e-9; Figure 8e). Model performance when predicting velocity was also improved (AICc: E[R]+History -36774.7, E[R] -36742.4). A detailed view is also given in Figure S11E, showing that this effect was not merely driven by the increasing reward history value estimates during the beginning of a block.

Models including both reward history and learned value best predicted vigor

Lastly, both model-derived terms, learned value and reward history, were included into a single generalized regression model (Table 1). We then compared this model via likelihood ratio test to an alternative model that excluded the reward history term. We found a significant improvement in model likelihood when reward history was included compared to its exclusion (LRT: χ_1^2 =30.156, p=3.987e-8). Interestingly, this model contradicted previous models that had indicated significant effects of trial number on peak velocity. It can be said, then, that the ongoing learning of value, and effect of relative reward context, could explain the trial-dependency of velocity. This result strengthens our argument that both recent, target-agnostic rewards and effort-conscious valuation of specific targets influence the motivation of movement. This effect of target-agnostic reward history is not explained by current trial value estimate, visualized in figure 8f. With all else being equal, outgoing peak velocity significantly varied with reward history. Model diagnostics confirmed that the

estimated reward history term was sufficiently orthogonal to other model terms, including current trial value estimate ($VIF_{History}$=1.12 [1.10,1.15 95% CI]). Model comparisons against previous regression models that included reward expectation, as well as the reward history term, reveal that the model including both value and history had the smallest AIC score, and thus best describes potential future data (AICc: Value+History -36880.5, Value -36852.3, E[R]+History -36774.7, E[R] -36742.4).

Table 1. Estimated regression parameters, confidence intervals, and Satterthwaite-approximated P-values for linear mixed model predicting the logarithm of outgoing peak velocity, incorporating terms for both reward history and estimated learned value, derived from a delta-rule model. Estimated random intercept variance $\sigma_{Participant}$ was 0.1347.

	Coefficient Estimate	95% confidence interval	P-value
(Intercept)	-1.1221	-1.2760 – -0.9682	**<0.0001**
Target [2]	-0.0622	-0.0694 – -0.0549	**<0.0001**
Target [3]	-0.0310	-0.0379 – -0.0242	**<0.0001**
Target [4]	-0.0507	-0.0580 – -0.0434	**<0.0001**
Block	0.1714	0.0747 – 0.2680	**0.0005**
Learned Value	0.0238	0.0018 – 0.0457	**0.0338**
Trial In Block	0.0134	-0.0340 – 0.0609	0.5786
Reward History	0.0831	0.0533 – 0.1130	**<0.0001**
Target Direction Repeat [True]	0.0255	0.0199 – 0.0312	**<0.0001**
Learned Value × Trial In Block	-0.0009	-0.0327 – 0.0309	0.9543

DISCUSSION

Here, we demonstrated that reach vigor tracks canonical variables of learning and motivation across time scales ranging from milliseconds to minutes. Velocity was modulated by reward expectation, reward prediction error and reward rate, key variables that have also been associated with striatal dopaminergic fluctuations. These results point to a potential neural mechanism by which dopamine can provide the bridge between decision making and movement control.

Reward prediction error is typically examined with respect to recent reward history and current target expectation. Higher value relative to history provides invigoration, and lesser value leads to comparative enervation. In this study, we investigate not only the effect of recent reward history on human arm reaching, but also the within-trial prediction error that results from stochastic reward experiences. Critically, we show that this prediction error can influence movements in an online fashion, dynamically increasing or decreasing relative velocity proportionally to the magnitude of the prediction error.

It may seem unlikely that the absence or presence of the reward feedback visual stimulus may influence ongoing velocity at sub-second speeds, approximately 215 ms after stimuli presentation. However, previous experimental evidence has shown that neurons in the superior colliculus, that then project to dopaminergic neurons within the VTA or SNc, have a response latency between 40 and 60 ms[34]. Another study, this time in rhesus monkeys, found median response latencies of cue-elicited DAN responses within the SNc were 112 ms (iqr of 92 – 172 ms)[35]. Others still have found that

modulation of rapid motor responses, on the order of 150 ms, is dependent on task specific visuospatial features[36]. Reward prediction error modulation in saccade latencies were found to be affected on the order of ~150 ms[37]. A more recent study has found that changes in direction of movement, specifically when a reach target suddenly decreases in expected value, occur on average ~248 ms after cue[38]. This body of literature detailing the short-latency rapid motor response suggests that the rapid response in relative velocity that we see can be attributed to tuned sensorimotor reward-based predictions.

Beyond the effects of within-trial reward and prediction error, we investigated the relationship between biomechanical effort, subjective choice preference, and movement vigor. Aside from reward-mediated invigorative effects of DA, both in in movement and deliberation, other research has found the neurotransmitter instrumental in overcoming costs and providing motivation in the face of effortful action[39–43], which has been shown to influence motor behavior and deliberation [44–46]. However, phasic midbrain DAN activity may not necessarily code for upcoming effort value[47]. In monkey (Macaca mulatta) SNc, only a minority 13% of spiking activity in a reaching task was correlated with net utility (rewards minus cost) compared to reward-alone. One hypothesis is that the phasic DA response incorporates effort costs only if reward rate, or discounted value of future action, were meaningfully affected[43]. Other work, however, has shown prediction error responses in Japanese macaque substantia nigra to be significantly influenced by relative effort of a saccade-to-fixation task[48,49]. Paradoxically, RPE signals were enhanced by experienced effort costs, perhaps mediated by the incurred effort expenditure either by increasing the reward signal or decreasing the predicted reward signal. Regardless, response at the time of cost-cue presentation was diminished for high-cost compared to low-cost efforts. Other neurotransmitters and pathways may be implicated in utility prediction. Specific serotonin receptor activation in mice was found to increase motivation and vigor alongside an increase in dorsomedial striatum extracellular DA[50]. Thus, the effect of variable effort on the behavioral and dopaminergic response to reward prediction error remains an open question.

The use of a delta-rule learning model to estimate and quantify the trial-to-trial, subjective value of a given target was motivated in part by its success in literature modelling prediction error responses in human dopaminergic systems[16,51–54]. Beyond these reward prediction effects, effort too has been shown to induce variable signals in rat ventral striatum and dorsal anterior cingulate cortex (ACCd)[55,56]. Likewise, human ACCd activity was found to reflect the interaction between both expected reward and expected effort[57]. We also modeled effort as additively discounting experienced reward, rather than multiplicatively as in previous work[58]. Additive discounting accounts for the invigorating effect of reward and aligns with the conceptual framework that the brain modulates movement vigor so as to maximize reward rate[31,59]. Our approach is validated in our data where the value prediction error, which integrates learned reward and effort costs, has a significant effect on the within-trial change in velocity (Figure 7).

However, it is possible that the prediction error does not incorporate the experienced target effort, which would be modeled as follows:

$$\hat{R}'^{T}_{i} = \hat{R}^{T}_{i} + \eta_s\left(R_i - \hat{R}^{T}_{i}\right)$$

$$V^{T}_{i} = \hat{R}^{T}_{i} + e^{T}_{s}$$

In other words, learning is driven by reward prediction, and behavior is a result of both learned reward and already known effort. Both models were fit to choice data, and we compared the two with respect to performance in predicting the participant population decisions. The first model, with learning driven from a value prediction error, had greater conditional R^2, 0.686, compared to the alternative (0.543). However, AICc scores show a preference towards model 2: 2890.61 compared to 2917.76. From our experimental design, the kinematic data would be equally well described with either model, as effort values were unmanipulated and constant to each target direction.

Ultimately, the evidence is ambiguous to state conclusively which of the feasible delta-rule models for learning target value best describes the data. Our experiment did not originally set out to test these distinctions, and so additional follow-up with the requisite design considerations would be needed. We can say, however, to speak to our original hypothesis that learning does occur, and a prediction error does drive changes in reach kinematics on the order of hundreds of milliseconds.

Our experiment also highlights the many differences between an experienced stochasticity and described one[60], although our first experiment may be better described as a *hybrid* protocol, as trial-to-trial feedback is still provided to reinforce the previously stated expected reward frequencies at the beginning of the block. The use of kinematic responses, specifically peak velocity on outgoing and return portions of the movement, are useful tools in understanding not just these differences[61], but also the similarities between the experienced and described rewarding environments. Online reward prediction error was remarkably similar, occurring within the same time frame and showing comparable magnitudes of response. The most apparent distinction, however, was the influence of recent reward history. In our second experiment, the found relationship between prior reward and velocity can be said to have a longer view of the past, integrating incrementally over many trials to arrive in near proximity to the underlying arithmetic average. The first experiment could be said to elicit more impulsive or rapid response, with the influence of trials further back in time quickly diminishing. This behavior may speak directly to the difference in environmental uncertainties, where average reward expectation is immediately known to the participant compared to when it must be experienced and learned over time. Previous research has found significant effects of environmental uncertainty in mice and monkey reward learning rates[62], with greater uncertainty resulting in reducing update rates, which is reflected in the results presented here in human vigor response and the difference in reward history α-coefficients between the two experiments. It too may explain the differences in the change in average reaction time responses over the course of the experiment. In our first protocol, with known reward frequencies, reaction time decreased from block to block. In the second, the opposite effect was observed where average reaction times, specifically within the single-target trials, increased from block to block.

Conclusion: In the two experiments presented, we demonstrate the sensitivity of movement vigor not only to reward, but its probability of reception. As the likelihood of reward feedback increased, so too did vigor. Change in vigor on the return portion of the movement was also dependent on the reward prediction error, the difference between experienced feedback and its expectation. Lastly, in our second experiment, we see that trial-specific learned value, modelled via a delta-rule update formulation, could incorporate both rewards and subjective efforts to predict individual reaching vigor. Target-agnostic reward history, in both experiments, was found to significantly influence vigor, even when the value expectation was matched.

MATERIALS AND METHODS

Participants: The study consisted of two experiments, each with an independent population (first: n=42, f=16, age=22.3±0.76; second: n=22, f=16, age=23.5±0.9). All individuals were either ambidextrous or right-handed as determined by the Edinburgh Handedness Inventory survey, as well as free of upper extremity injury or self-reported neurological condition. In the first experiment, participants were compensated at a fixed rate of $10 per hour with no difference depending on performance. For the second experiment, the $10/hr rate remained with the potential of an additional $5 depending on performance (average compensation = $14.02±0.13). Participants provided written and informed consent, and procedures were approved of by the University of Colorado Boulder institutional review board.

Experimental Design: Participants performed a bandit-like task using the KINARM end-point robotic arm (BKIN Technologies, ON, CA) to control a white cursor presented onscreen. From a home circle (r=1cm) at the center of a ring of radius 10 cm, individuals were instructed to make an out-and-back reaching motion to move a cursor (r=0.5cm) a cued target (r=1cm), displayed at 45, 135, 225, or 315 degrees relative to the home circle. Target cues were displayed at a variable time interval uniformly ranging from 800 ms to 1000 ms after trial initiation. On reaching to the cued target, participants need not hit the prompted target exactly. If absolute angular error at the 10 cm radial distance was less than 22°, the target was counted as "hit." If individuals failed to hit the prompted target, either within the ±22° accuracy constraint or within 4 seconds, a large red "X" appeared onscreen with an accompanying tone (400 Hz, 100 ms duration) indicating a trial failure.

Individuals were instructed that each target had a unique probability (100%, 66%, 33%, or 0%) of providing rewarding feedback at the moment of target hit. Reward feedback consisted of a brief, high-pitched tone (880 Hz, 100 ms duration) being played, the target instantly doubling in size, changing color to yellow, and blinking intermittently (2 blinks, on-off duration at 50 ms). Sample trial progression is shown in figure 1c. Reward probability per target changed between blocks, with a total of 4 blocks each consisting of 180 trials. Trial order and rewards were pseudorandomized such that for each set of 36 trials, each target was cued 9 times, and reward feedback per target was given 9, 6, 3 or 0 times for the reward expectations of 100%, 66%, 33%, and 0% respectively. Individuals were not told the total number of trials to be experienced, only that the total time in the experiment would require approximately 1 hour.

Reward prediction error was defined as the difference between the received feedback and the presented target's reward expectation, i.e.:

$$R_i \in \{0,1\}$$

$$RPE_i = R_i - E[R_i^T]$$

Thus, the four target reward probabilities led to five different reward prediction errors varying in both magnitude and sign. The probabilistic targets (33% and 66%) led to both smaller and larger, positive and negative reward prediction errors. The deterministic targets (0% and 100%) led to 0 RPE, but also allowed us to compare the effects of reward feedback per se, independent of reward prediction error.

In experiment 1, participants were explicitly told which targets were to have what expectation of reward within a block. In experiment 2, the rewards were identical to experiment 1 except that

participants were not told of the target reward probabilities but rather had to learn from experience when reaching to the target when prompted. To quantify the degree to which they learned the reward probabilities, the final 36 trials of each block were 'choice trials' where two targets were cued rather than a single target. No reward feedback was provided within these choice trials after a target was hit on the reach. Individuals were familiarized with the nature of these choice trials beforehand and were informed that the underlying reward frequencies remained constant within a block. Each unique choice pair (n=6) was presented six times during these choice trial periods. For choice trials, we analyzed the choices as a function of expected value and expected reward. We also measured the response time, the time between target presentation and movement onset. Participants were instructed in the second experiment that additional compensation was contingent on the potential rewards received during these choice trials and thus were incentivized to reach towards targets with previously experienced greater expectation for reward feedback. Total bonus compensation was calculated as the sum of chosen expected reward, divided by maximum potential expected chosen reward (18.66), times five dollars:

$$\frac{\sum_i E[R]_{choice,i}}{18.66} \times \$5$$

In the first experiment, individuals were familiarized with the reaching task by completing 16 out-and-back reaches, four each to the four different targets. In the second experiment, an additional 8 choice trials were provided as familiarization. Reward feedback was withheld on all familiarization trials.

Kinematic Metrics: KINARM encoders sampled hand position and velocity (x,y cartesian coordinates) at 1000 Hz. Raw kinematics data were filtered via third-order double pass filter with cutoff of 10 Hz. Radial velocity was computed from numerical differentiation of the radial position relative to the home circle using a second order centered finite difference. The primary metrics on each trial (both single target trials and choice trials) were movement initiation time (reaction time), peak outgoing and return radial velocities, maximum excursion radial distance, and movement duration from onset to target hit. Reaction time, defined as the time between cue presentation and detected movement onset, was calculated by use of the MACC-based onset detection method[63].

Trial Exclusion Criteria: For the first experiment, 1.83±0.35% of trials per participant on average were excluded from kinematic analysis either due to failure to complete the trial, failing to reach the outer target ring during the initial outward reach (a “double-peak” movement), or missing the target completely (angular error was ≥22.5°). In the second experiment 7±2.37% of single-target trials per participant on average were excluded from further kinematic analysis. These trials were included, however, for the purposes of calculating reward history and recorded as providing no reward feedback.

Linear Regression Models: To determine the relationship between velocity, target reward, and reward prediction error, we used generalized Gamma linear mixed models with a log link function to estimate the relative effects of factors of interest and to account for between-participant variability. The Gamma distribution was selected after post-hoc residual analysis of a gaussian-residual assumption revealed significant heteroskedasticity and skewness. The log-link was ultimately selected after model comparisons to different potential link functions (inverse and identity). A similar analysis was used to probe the relationship between reaction time and target reward expectation.

In the first experiment, our regression models for peak velocity and reaction time was as follows:

$$\log(\mu_y) = (1 + \text{Target} * \text{Repeat} + \text{Block} * \text{Trial} + \text{E[R]} + \text{Prior}_{\text{RWD}}) + (1 + \text{E[R]} + \text{Target} + \text{Block}|Subj)$$

With μ_y representing the mean of the outcome of interest, i.e., velocity or reaction time.

And for the second experiment, an additional interaction term with trial was added:

$$\log(\mu_y) = (1 + \text{Target} * \text{Repeat} + \text{Block} * \text{Trial} + \text{E[R]} + \text{Prior}_{\text{RWD}} + \text{E[R]}\text{:}\,\text{Trial}) + (1 + \text{E[R]} + \text{Target} + \text{Block}|Subj)$$

Target is treated as factored variable, and all others as continuous variables. The regression model for return peak velocity was similar, but with added terms for reward reception and prediction error while controlling for outward peak velocity:

$$\log(\mu_y) = (1 + \text{OutPV} * \text{Target} + \text{Repeat} + \text{Trial} + \text{RPE} * \text{Reward}) + (1 + \text{RPE} * \text{Reward} + \text{OutPV} + \text{Trial}|Subj)$$

For velocity difference models, a gaussian distribution was found to better approximate the residuals, so a typical LMER with an identity link function was used instead of the generalized Gamma model. All mixed regression models were fitted to restricted maximum likelihood (REML) with the *lme4* and *lmerTest* packages for the *R* language[64,65], except for when model comparisons were performed with the *performance* R package[66]. P-values for mixed model regression coefficients are derived from Satterthwaite approximations. Hypothesis testing for linear combinations (LHT) of regression coefficients and calculation of asymptotic Chi-squared test statistics was performed with the linearHypothesis function in the *car* R package[67].

One-Dimensional SPM Analysis: Continuous time analysis of radial velocity was performed by one-dimensional statistical parametric mapping[68]. Biomechanical curves are well-suited to analysis by this methodology, owing to their smoothness and discrete bounds. We measured the potential effect of reward feedback on different expectation contexts, either deterministic or stochastic, with either single-sample t-tests or 2-factor ANOVAs respectively. This was done to both test for potential significant effects as well as when such effect may occur on instantaneous radial velocity. We focused on the time window from 150 ms prior to hitting the target (i.e. feedback reception) and 300 ms afterward.

Population-level inference on the effect of reward prediction error on within-trial normalized velocity (taken as instantaneous velocity divided by the within-trial peak outgoing radial velocity) was conducted in a manner following previous work[69]. First, regression analyzes were fitted per participant, then this collection of 1-D beta values were submitted to a second-level single-sample t-test.

Learning Model Design: To model the trial-to-trial update process in experiment 2, we developed a Bayesian hierarchical Rescorla-Wagner learning model. In each of a block's single-target trials, the presented target's estimated value was updated based on this rule:

$$V'^{T}_{i} = V^{T}_{i} + \eta_s\left(R_i + e^{T}_{s} - V^{T}_{i}\right)$$

With *T* the presented target at 45, 135, 225, or 315 degrees, *R* the reward feedback [0,1], *e* the effort cost, *s*: participant, *i*: trial, and η: learning rate. Participant-specific learning rates (η_s) and effort costs

(e_s) were found via posterior maximum likelihood estimation based on choices. The full model (with priors) is characterized as follows:

$$(Choice_i = A) \sim \text{Bernoulli}\left(\frac{1}{1+\exp\left(-\tau_s\left(V^T_{A\ i} - V^T_{B\ i}\right)\right)}\right)$$

$$\bar{\zeta} \sim N(0,10);\ \bar{e} \sim N(0,5);\ \bar{\tau} \sim N(0,5)$$

$$\sigma_\zeta \sim \Gamma(2,0.01);\ \sigma_e \sim \Gamma(2,0.01);\ \sigma_\tau \sim \Gamma(2,0.01)$$

$$\zeta_s \sim N\left(\bar{\zeta}, \sigma_\zeta\right);\ e_s \sim N(\bar{e}, \sigma_e);\ \tau_s \sim N(\bar{\tau}, \sigma_\tau)$$

$$\eta_s = \frac{1}{1+\exp(-\zeta_s)}$$

In each choice trial, targets were randomly labelled as either A or B. Value prediction error was defined as the sum of reward feedback and experienced effort minus learned value:

$$VPE_i = R_i + e_s^T - V_i^T$$

Sampling for Bayesian models was performed by NUTS Hamiltonian MCMC algorithm via use of the *RStan* package[70].

Reward History Model: To investigate the effect of prior reward on excursion peak velocity, we applied an exponential moving average to reward reception history[33]. For each single-target trial, average reward was updated with this rule:

$$\bar{R}_{i+1} = \alpha R_i + (1-\alpha)\bar{R}_i$$

The smoothing factor, α, was fit to maximize the likelihood of a reaction time random intercept regression[18]:

$$\log(\mu_{RT}) \sim (1 + \text{Block} + \bar{R} + \text{Repeat} + \text{Trial} * \text{E[R]} + \text{Target}) + (1|Subj)$$

The effects of block number (Block), direction repetition (Repeat), trial number (Trial), reward expectation (E[R]) and factored target direction (Target) were included as control variables. We then used this model-derived reward history term to independently predict outgoing peak velocity in subsequent linear models.

REFERENCES

1. Beierholm, U. *et al.* Dopamine Modulates Reward-Related Vigor. *Neuropsychopharmacology* **38**, 1495–1503 (2013).
2. Choi, J. E. S., Vaswani, P. A. & Shadmehr, R. Vigor of Movements and the Cost of Time in Decision Making. *J Neurosci* **34**, 1212–1223 (2014).
3. Opris, I., Lebedev, M. & Nelson, R. J. Motor Planning under Unpredictable Reward: Modulations of Movement Vigor and Primate Striatum Activity. *Front Neurosci* **5**, (2011).
4. Reppert, T. R., Lempert, K. M., Glimcher, P. W. & Shadmehr, R. Modulation of Saccade Vigor during Value-Based Decision Making. *Journal of Neuroscience* **35**, 15369–15378 (2015).
5. Shadmehr, R. & Ahmed, A. A. *Vigor: Neuroeconomics of Movement Control*. (MIT Press, 2020).
6. Summerside, E. M., Shadmehr, R. & Ahmed, A. A. Vigor of reaching movements: reward discounts the cost of effort. *Journal of Neurophysiology* **119**, 2347–2357 (2018).
7. Smoulder, A. L. *et al.* A neural basis of choking under pressure. *Neuron* **112**, 3424-3433.e8 (2024).
8. Korbisch, C. C., Apuan, D. R., Shadmehr, R. & Ahmed, A. A. Saccade vigor reflects the rise of decision variables during deliberation. *Current Biology* **32**, 5374-5381.e4 (2022).
9. Ungerstedt, U. Adipsia and Aphagia after 6-Hydroxydopamine Induced Degeneration of the Nigro-striatal Dopamine System. *Acta Physiologica Scandinavica* **82**, 95–122 (1971).
10. Klaus, A., Alves da Silva, J. & Costa, R. M. What, If, and When to Move: Basal Ganglia Circuits and Self-Paced Action Initiation. *Annu. Rev. Neurosci.* **42**, 459–483 (2019).
11. Larry, N., Zur, G. & Joshua, M. Organization of reward and movement signals in the basal ganglia and cerebellum. *Nat Commun* **15**, 2119 (2024).
12. Dudman, J. T. & Krakauer, J. W. The basal ganglia: from motor commands to the control of vigor. *Current Opinion in Neurobiology* **37**, 158–166 (2016).

13. Thura, D. & Cisek, P. The Basal Ganglia Do Not Select Reach Targets but Control the Urgency of Commitment. *Neuron* **95**, 1160-1170.e5 (2017).
14. Herz, D. M. *et al.* Dynamic control of decision and movement speed in the human basal ganglia. *Nat Commun* **13**, 7530 (2022).
15. Schultz, W. Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology* **80**, 1–27 (1998).
16. Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299**, 1898–1902 (2003).
17. McDougle, S. D. *et al.* Neural Signatures of Prediction Errors in a Decision-Making Task Are Modulated by Action Execution Failures. *Current Biology* **29**, 1606-1613.e5 (2019).
18. Hamid, A. A. *et al.* Mesolimbic dopamine signals the value of work. *Nat Neurosci* **19**, 117–126 (2016).
19. Niv, Y., Daw, N. D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* **191**, 507–520 (2007).
20. Wang, Y., Toyoshima, O., Kunimatsu, J., Yamada, H. & Matsumoto, M. Tonic firing mode of midbrain dopamine neurons continuously tracks reward values changing moment-by-moment. *eLife* **10**, e63166 (2021).
21. Ikemoto, S. & Panksepp, J. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Res Brain Res Rev* **31**, 6–41 (1999).
22. Salamone, J. D. & Correa, M. Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behav Brain Res* **137**, 3–25 (2002).

23. Cai, X. *et al.* Dopamine dynamics are dispensable for movement but promote reward responses. *Nature* 1–9 (2024) doi:10.1038/s41586-024-08038-z.

24. Engel, L. *et al.* Dopamine neurons drive spatiotemporally heterogeneous striatal dopamine signals during learning. *Current Biology* **34**, 3086-3101.e4 (2024).

25. Howe, M. W. & Dombeck, D. A. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* **535**, 505–510 (2016).

26. Mohebi, A., Wei, W., Pelattini, L., Kim, K. & Berke, J. D. Dopamine transients follow a striatal gradient of reward time horizons. *Nat Neurosci* 1–10 (2024) doi:10.1038/s41593-023-01566-3.

27. Panigrahi, B. *et al.* Dopamine Is Required for the Neural Representation and Control of Movement Vigor. *Cell* **162**, 1418–1430 (2015).

28. da Silva, J. A., Tecuapetla, F., Paixão, V. & Costa, R. M. Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* **554**, 244–248 (2018).

29. Syed, E. C. J. *et al.* Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat Neurosci* **19**, 34–36 (2016).

30. Schall, T. A. *et al.* Temporal dynamics of nucleus accumbens neurons in male mice during reward seeking. *Nat Commun* **15**, 9285 (2024).

31. Shadmehr, R., Huang, H. J. & Ahmed, A. A. A Representation of Effort in Decision-Making and Motor Control. *Current Biology* **26**, 1929–1934 (2016).

32. Wagner, A. R. & Rescorla, R. A. Inhibition in Pavlovian conditioning: Application of a theory. *Inhibition and learning* 301–336 (1972).

33. Daw, N. D., Kakade, S. & Dayan, P. Opponent interactions between serotonin and dopamine. *Neural Networks* **15**, 603–616 (2002).

34. Dommett, E. *et al.* How Visual Stimuli Activate Dopaminergic Neurons at Short Latency. *Science* **307**, 1476–1479 (2005).

780 35. Ravel, S. & Richmond, B. J. Dopamine neuronal responses in monkeys performing visually cued
781 reward schedules. *European Journal of Neuroscience* **24**, 277–290 (2006).
782 36. Pruszynski, J. A. *et al.* Stimulus-locked responses on human arm muscles reveal a rapid neural
783 pathway linking visual input to arm motor output. *European Journal of Neuroscience* **32**, 1049–
784 1057 (2010).
785 37. Sedaghat-Nejad, E., Herzfeld, D. J. & Shadmehr, R. Reward Prediction Error Modulates Saccade
786 Vigor. *J. Neurosci.* **39**, 5010–5017 (2019).
787 38. Marti-Marca, A., Deco, G. & Cos, I. Visual-reward driven changes of movement during action
788 execution. *Sci Rep* **10**, 15527 (2020).
789 39. Collins, A. G. E. & Frank, M. J. Surprise! Dopamine signals mix action, value and error. *Nat*
790 *Neurosci* **19**, 3–5 (2016).
791 40. Hagura, N., Haggard, P. & Diedrichsen, J. Perceptual decisions are biased by the cost to act.
792 *eLife* **6**, e18422 (2017).
793 41. Kurniawan, I., Guitart-Masip, M. & Dolan, R. Dopamine and Effort-Based Decision Making.
794 *Frontiers in Neuroscience* **5**, (2011).
795 42. Salamone, J., Correa, M., Farrar, A., Nunes, E. & Pardo, M. Dopamine, behavioral economics,
796 and effort. *Frontiers in Behavioral Neuroscience* **3**, (2009).
797 43. Walton, M. E. & Bouret, S. What Is the Relationship between Dopamine and Effort? *Trends in*
798 *Neurosciences* **42**, 79–91 (2019).
799 44. Courter, R. J., Alvarez, E., Enoka, R. M. & Ahmed, A. A. Metabolic costs of walking and arm
800 reaching in persons with mild multiple sclerosis. *Journal of Neurophysiology* **129**, 819–832
801 (2023).
802 45. Sukumar, S., Shadmehr, R. & Ahmed, A. A. Effects of reward and effort history on decision
803 making and movement vigor during foraging. *Journal of Neurophysiology* **131**, 638–651 (2024).

804 46. Summerside, E. M. & Ahmed, A. A. Using metabolic energy to quantify the subjective value of
805 physical effort. *Journal of The Royal Society Interface* **18**, 20210387 (2021).
806 47. Pasquereau, B. & Turner, R. S. Limited encoding of effort by dopamine neurons in a cost-benefit
807 trade-off task. *J Neurosci* **33**, 8288–8300 (2013).
808 48. Tanaka, S., O'Doherty, J. P. & Sakagami, M. The cost of obtaining rewards enhances the reward
809 prediction error signal of midbrain dopamine neurons. *Nat Commun* **10**, 3674 (2019).
810 49. Tanaka, S., Taylor, J. E. & Sakagami, M. The effect of effort on reward prediction error signals in
811 midbrain dopamine neurons. *Current Opinion in Behavioral Sciences* **41**, 152–159 (2021).
812 50. Bailey, M. R. *et al.* An Interaction between Serotonin Receptor Signaling and Dopamine
813 Enhances Goal-Directed Vigor and Persistence in Mice. *J. Neurosci.* **38**, 2149–2162 (2018).
814 51. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for
815 exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
816 52. Maes, E. J. P. *et al.* Causal evidence supporting the proposal that dopamine transients function
817 as temporal difference prediction errors. *Nat Neurosci* **23**, 176–178 (2020).
818 53. Starkweather, C. K. & Uchida, N. Dopamine signals as temporal difference errors: recent
819 advances. *Current Opinion in Neurobiology* **67**, 95–105 (2021).
820 54. Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning.
821 *Nat Neurosci* **16**, 966–973 (2013).
822 55. Schweimer, J. & Hauber, W. Involvement of the rat anterior cingulate cortex in control of
823 instrumental responses guided by reward expectancy. *Learn. Mem.* **12**, 334–342 (2005).
824 56. Walton, M. E., Bannerman, D. M., Alterescu, K. & Rushworth, M. F. S. Functional Specialization
825 within Medial Frontal Cortex of the Anterior Cingulate for Evaluating Effort-Related Decisions. *J.*
826 *Neurosci.* **23**, 6475–6479 (2003).

827 57. Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E. J. & Rushworth, M. F. S. Effort-Based
828 Cost–Benefit Valuation and the Human Brain. *J. Neurosci.* **29**, 4531–4541 (2009).
829 58. Klein-Flügge, M. C., Kennerley, S. W., Saraiva, A. C., Penny, W. D. & Bestmann, S. Behavioral
830 Modeling of Human Choices Reveals Dissociable Effects of Physical Effort and Temporal Delay
831 on Reward Devaluation. *PLOS Computational Biology* **11**, e1004116 (2015).
832 59. Emlen, J. M. The Role of Time and Energy in Food Preference. *The American Naturalist* **100**, 611–
833 617 (1966).
834 60. Garcia, B., Cerrotti, F. & Palminteri, S. The description–experience gap: a challenge for the
835 neuroeconomics of decision-making under uncertainty. *Philosophical Transactions of the Royal*
836 *Society B: Biological Sciences* **376**, 20190665 (2021).
837 61. Fitzgerald, T. H. B., Seymour, B., Bach, D. R. & Dolan, R. J. Differentiable neural substrates for
838 learned and described value and risk. *Curr Biol* **20**, 1823–1829 (2010).
839 62. Woo, J. H. *et al.* Mechanisms of adjustments to different types of uncertainty in the reward
840 environment across mice and monkeys. *Cogn Affect Behav Neurosci* **23**, 600–619 (2023).
841 63. Botzer, L. & Karniel, A. A simple and accurate onset detection method for a measured bell-
842 shaped speed profile. *Frontiers in Neuroscience* **3**, (2009).
843 64. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4.
844 *Journal of Statistical Software* **67**, 1–48 (2015).
845 65. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed
846 Effects Models. *Journal of Statistical Software* **82**, 1–26 (2017).
847 66. Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P. & Makowski, D. performance: An R
848 Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source*
849 *Software* **6**, 3139 (2021).
850 67. Fox, J. & Weisberg, S. *An R Companion to Applied Regression*. (Sage, Thousand Oaks CA, 2019).

68. Pataky, T. C. Generalized *n*-dimensional biomechanical field analysis using statistical parametric mapping. *Journal of Biomechanics* **43**, 1976–1982 (2010).

69. Beckmann, C. F., Jenkinson, M. & Smith, S. M. General multilevel linear modeling for group analysis in FMRI. *NeuroImage* **20**, 1052–1063 (2003).

70. Stan Development Team. RStan: the R interface to Stan. (2024).