Article

Current Biology

An Aha moment precedes the strategic response to a visuomotor rotation

Highlights

- Humans often perseverate in their baseline aiming strategy for many trials
- The switch from perseveration to a new strategy is discrete, not gradual
- The first new strategy is often correct in magnitude
- Uncertainty about the direction of the perturbation persists through early learning

Authors

Max Townsend, Matthew Warburton, Carlo Campagnoli, Mark Mon-Williams, Faisal Mushtaq, J. Ryan Morehead

Correspondence

max.o.b.townsend@gmail.com

In brief

Townsend et al. discover that strategic compensation for a sensorimotor perturbation emerges from an "Aha!" moment, producing a discrete jump to (approximately) the right solution. They demonstrate, through computational modeling, that this offers a better account of participant behavior than gradual error minimization or trial-and-error learning.

Townsend et al., 2026, Current Biology *36*, 1–13
May 18, 2026
https://doi.org/10.1016/j.cub.2026.04.021

CellPress

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

Current Biology

CellPress
OPEN ACCESS

Article

An Aha moment precedes the strategic response to a visuomotor rotation

Max Townsend,[1,5,*] Matthew Warburton,[1] Carlo Campagnoli,[1,2] Mark Mon-Williams,[1,3,4] Faisal Mushtaq,[1,2,4] and J. Ryan Morehead[1]

[1]School of Psychology, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK
[2]Centre for Immersive Technologies, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK
[3]Bradford Institute for Health Research, Bradford Hospitals National Health Service Trust, Duckworth Lane, Bradford BD9 6RJ, UK
[4]Leeds NIHR Biomedical Research Centre, Chapeltown Road, Leeds LS7 4SA, UK
[5]Lead contact
*Correspondence: max.o.b.townsend@gmail.com
https://doi.org/10.1016/j.cub.2026.04.021

SUMMARY

Strategic behavior in sensorimotor adaptation tasks is typically described as either a gradual error-minimization process or as a process of learning through trial and error. The former predicts a gradual monotonic reduction in error, until some asymptote, while the latter predicts behavioral exploration to discover an efficacious solution. Another explanation is that a sufficiently rich understanding of the task culminates in an "Aha!" moment and that this richer understanding naturally entails a new strategy. This predicts some period of perseveration in baseline behavior, followed by an abrupt shift to a new strategic solution. To avoid obfuscation caused by the motor system, we first investigated these hypotheses in a strategy-only aiming game, where participants aim and fire a cannon at targets. Most participants exhibited a period of baseline perseveration followed by a single-trial shift to a distinct strategy. This strategy typically appears to produce the correct compensatory magnitude for a noisily estimated perturbation, as well as substantial variability in compensatory sign. We then applied the same analyses to reaching data from a visuomotor rotation task that inhibited implicit adaptation through delaying feedback presentation. Similarly, we found that participants typically perseverated in reaching toward the target before suddenly switching, in a single trial, to good performance in compensatory magnitude. We fit our generative "Aha!" model to 1,337 participants across several datasets and show that it faithfully captures these phenomena. Our findings highlight the need to update existing models to account for the abrupt, insight-driven shifts that characterize individual human responses to sensorimotor perturbations.

INTRODUCTION

Correcting for movement errors is integral to the success of an embodied agent. Humans can strategically modify their behavior to expedite this error-correcting process.[1–9] While recent progress has been made in enumerating core strategic phenomena and in formulating frameworks from which to investigate these phenomena,[10] our understanding of strategic error correction remains shallow.

Traditionally, strategic behavior in visuomotor rotation (VMR) tasks has been characterized as a gradual error-reduction process. This view predicts a gradual, monotonic increase in learning that tends toward some asymptote, typically modeled by a state-space model (SSM).[3,8,11–13] While consistent with group-averaged learning curves, it is unclear if the individual time course of learning evolves in the same manner. Indeed, this gradual monotonic group curve can result from averaging together individual learning trajectories, each of which following distinct underlying dynamics.[14,15]

Thus, we should also consider that strategic behavior may instead reflect trial-and-error learning, where actions are selected based on their expected value or inferred efficacy. Here, the agent explores action space (or hypothesis space) to discover an efficacious solution. In this view, the evolution of behavior is either the result of a reinforcement learning (RL) process or of updating a set of hypotheses about the context, from which the efficacy of actions is inferred.[8,10,11,16–24] Learning in this manner typically generates clear signals of behavioral exploration before settling on a strategy. These behavioral patterns are inconsistent with, and dissociable from, gradual error reduction.

Importantly, the RL and hypothesis-testing accounts are both incomplete. They assume that the agent's initial task representation entails a strategy space that is conducive to discovering efficacious strategies. Therefore, we propose that a key component of strategic learning, ignored by extant theories, is the initial process of changing one's representation of the task and how one's goals relate to the task. Moreover, this initial updating of the task representation may naturally entail an efficacious strategy and eat much of the lunch of the trial-and-error accounts.

Indeed, there is much research on the "Aha!" moment. This "Aha!" moment may reflect a gain of insight—a qualitatively richer understanding of the task and its demands—that proceeds from reorganizing one's internal representation of the problem.[25–30] This kind of sudden learning has been observed

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

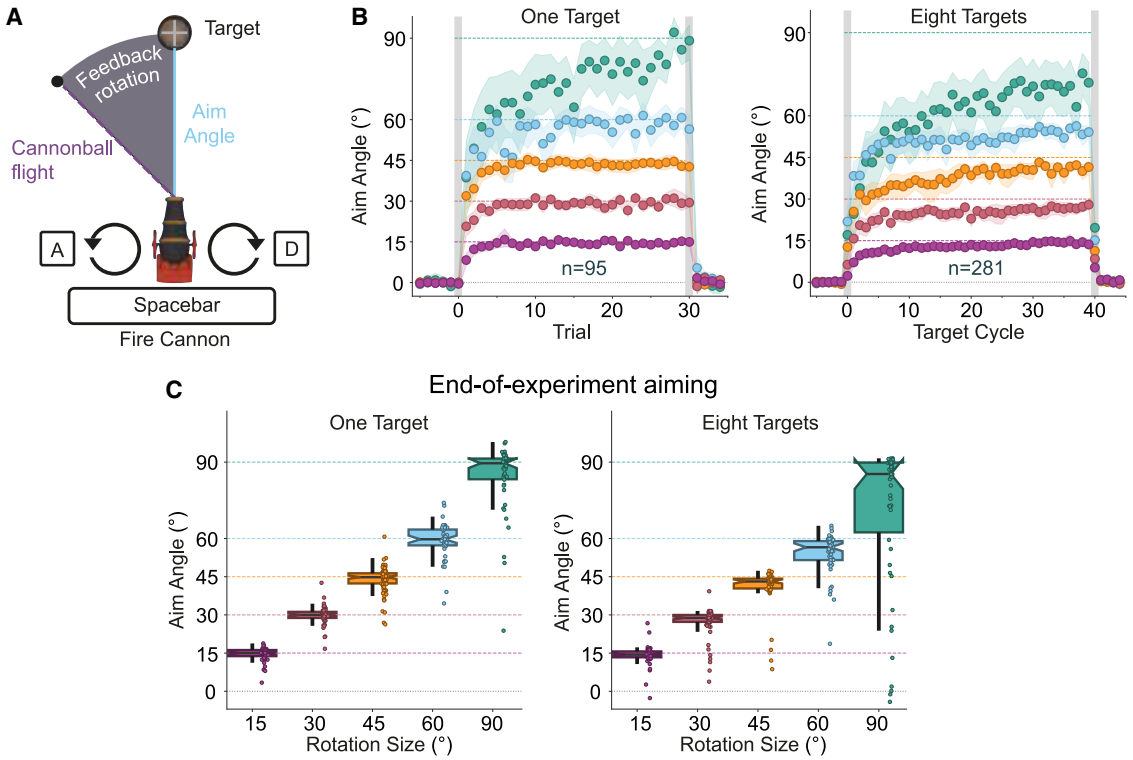


Figure 1. Strategic responses to VMRs were isolated in a canon aiming game

(A) The participant must aim the cannon to shoot at a target. The cannon is rotated using the "A" and "D" keys and fired using the "spacebar" key.

(B) Aiming behavior over the course of experiments 1a, with one target location (left), and 1b, with eight target locations (right). Points show group means, error bands show 95% CIs, and dashed lines indicate ideal aim angles.

(C) Aim angles during the final trial or cycle for experiments 1a (left) and 1b (right). The boxplots show the quartiles of the data, the whiskers bars show 1.5× the interquartile range, and the points show individual data.

in many areas of cognitive problem solving, such as the compound-word remote associations task,[31] anagrams,[32] rebus puzzles,[33] and the nine-dot problem and its variations,[34] among others.[35,36] Furthermore, we know from work by Manley et al.[37] that awareness of critical task properties may be required in learning to change reach directions, and work on consciousness suggests that this awareness may be all or none, rather than gradual.[38]

The strategic problem presented during the baseline (veridical) phase of a VMR task may be simply formulated as the requirement to aim at the target before firing. A change in the action-feedback mapping renders this parsimonious and effective representation uninformative as to how to act successfully. By reorganizing this representation to embody more semantic and relational information, an appropriate solution can be brought psychologically closer.[39] For example, when the action-feedback mapping is modified, a representation of the task that encodes a relationship between the aimed direction and the feedback direction would more naturally lead to an efficacious aiming solution. Learning would manifest, behaviorally, as a period of perseveration of baseline behavior, followed by a sudden, substantial shift in behavior following the "Aha!" moment. Indeed, this idea has been raised before in the VMR literature.[9,10]

To date, it has remained difficult to elucidate the true nature of strategic behavior in VMR tasks as extant task designs tend to leave strategic behavior inextricably linked to implicit adaptation and other contaminants of the motor system. This is either because one measure is used to infer the output of two or more processes, where it is difficult to correctly assign credit to these underlying processes,[40] or because these processes interact in a way that we do not currently understand. To make matters worse, different methods of probing strategic behavior generate distinct patterns of results.[41–43] This is likely because each method changes the nature of the interaction between the strategic system and the motor system. A different approach is to first understand each subprocess in relative isolation, such that we can use the findings as priors when studying the whole process. We can already remove implicit learning from the strategic context by presenting consistent clamped feedback that is irrelevant to performance.[44] We aim to complement this by isolating strategic learning from the motor-learning context while enforcing the same strategic demands.

Here, we investigated aiming behavior in a series of cannon-shooting tasks where the participant had to overcome a rotation of the visual feedback (Figure 1A). We found that aiming behavior evolved in a manner that was consistent with an "Aha!" moment, not merely trial-and-error learning or gradual error minimization. This pattern of baseline perseveration, followed by a delayed single-trial shift to good performance in magnitude, is clear both visually and through formal model comparison. We quantitatively

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

CellPress
OPEN ACCESS

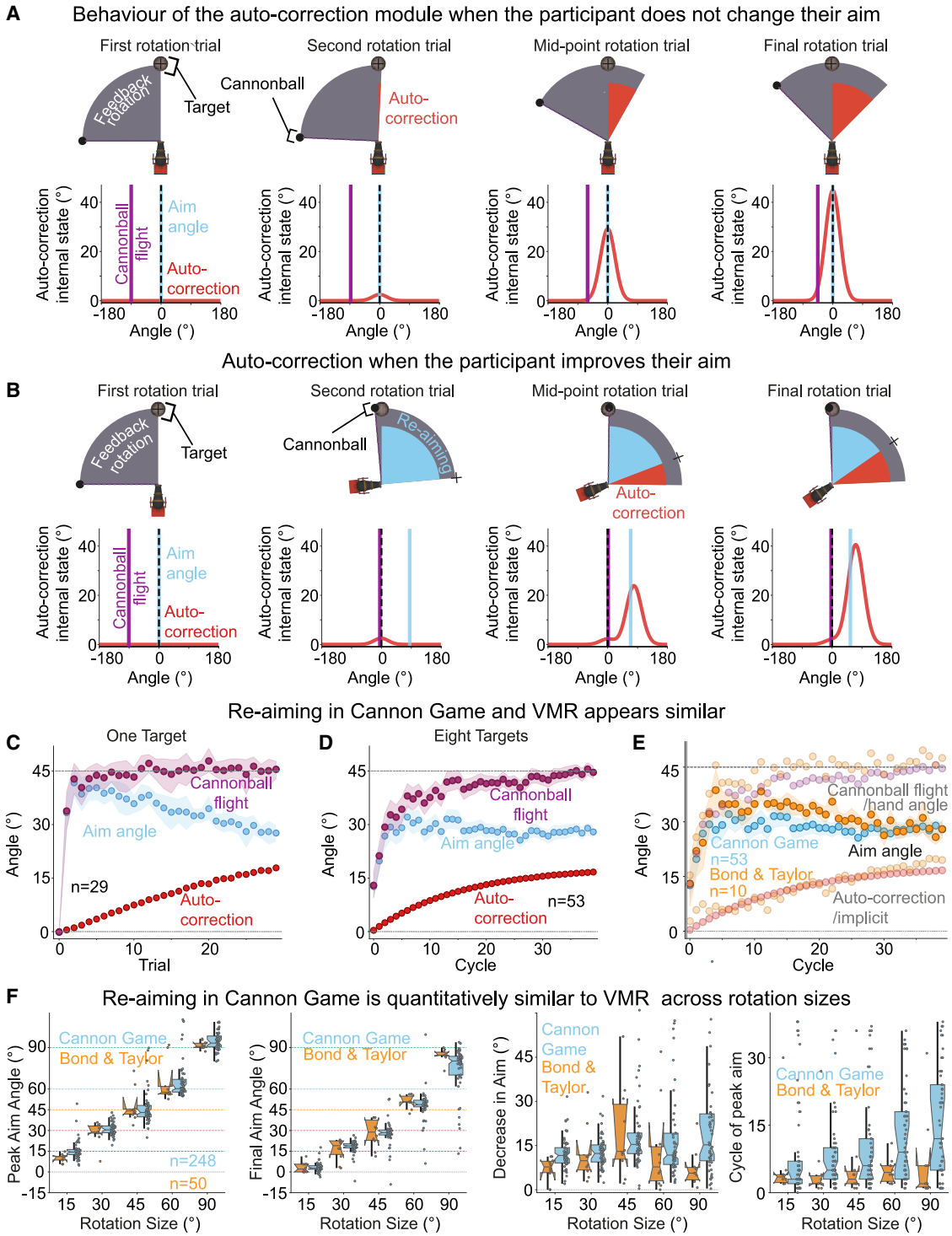


Figure 2. In experiments 2a and 2b, the cannon gradually learned to autocorrect for the feedback rotation

(A and B) The direction of cannonball flight (purple) is equal to autocorrection (red) + re-aiming (blue) − feedback rotation (gray). Time course of learning for the autocorrection module when a participant doesn't adjust his or her aim and time course of learning when a participant learns to re-aim, respectively. Top row: the contribution of each learning system (human re-aiming and cannon autocorrection) and the task outcome on a given trial. Bottom row: internal state of the autocorrection module (red) generalizes around the workspace.

(C and D) Performance in the 45° condition, under one-target (C) or eight-target (D) experiments.

(legend continued on next page)

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

confirmed that our task presented similar strategic demands as in a standard VMR task and then carried out the same modeling analyses on a previous dataset from a delayed-feedback VMR task[45] as well as on the comparator study itself.[1] We found the same pattern of results, where almost all participants were better described by an "Aha!" moment than by trial-and-error learning or gradual error minimization.

RESULTS

Experiment 1

We set out to characterize aiming behavior in a shooting task, where participants must aim and fire a cannon to hit targets (Figure 1A). This experiment removes the presence of any implicit motor learning (e.g., adaptation). In condition a, participants were exposed to either the ±15°, ±30°, and ±45° rotations or the ±45°, ±60°, and ±90° rotations at a single target. Participants quickly learned to compensate for all rotation sizes by aiming in the opposite direction (Figure 1B, left).

During the baseline phase, there was no bias in aiming (*t* tests against zero; Bonferroni-corrected $p > 0.334$). By the final five rotation trials of the block, for all rotation sizes, participants had learned to compensate for the feedback rotation (Figure 1C, left; *t* tests against zero; $p < 0.001$). By the final five rotation trials, the 90° group still aimed differently to the ideal aim angle (95% confidence interval [CI], 80.12°, 88.33°, $t(45) = 2.83$, $p = 0.034$), but aiming in all other groups was not different from the ideal (*t* tests against ideal; $p > 0.146$). For all rotations ($p > 0.467$), except 90° (95% CI, 62.75°, 81.98°, $t(45) = 3.69$, $p = 0.002$), this ideal aim angle was already achieved by the middle five trials. After the first washout trial, there was no residual bias in aiming (*t* test against zero; $p > 0.103$).

In condition b, five new groups of participants were each exposed to one of ±15°, ±30°, ±45°, ±60°, or ±90° rotations. With eight target locations, this condition comprised a single block presenting one rotation size. For all groups, there was no bias in baseline aiming ($p = 1$), and by the final five rotation cycles, aims were non-zero (Figure 1C; $p < 0.001$). During this late-rotation phase, aiming was smaller than ideal in the 30° group (95% CI, 25.4°, 28.65°, $t(57) = 3.67$, $p = 0.002$), 45° group (95% CI, 38.95°, 43.09°, $t(56) = 3.86$, $p = 0.001$), 60° group (95% CI, 52.17°, 56.36°, $t(56)$, $p < 0.001$), and the 90° group (95% CI, 61.7°, 88.32°, $t(53) = 5.48$, $p < 0.001$) but not in the 15° group ($p = 0.646$). During the middle five cycles, aiming in all groups was not ideal ($p < 0.002$). After the first washout cycle, there remained a small bias in aiming toward the previously ideal angle in the 15° (95% CI, 0.24°, 0.98°, $t(54) = 3.33$, $p = 0.007$) and 60° (95% CI, 0.12°, 0.54°, $t(56) = 3.1$, $p = 0.015$) groups but not in others ($p > 0.094$).

Experiment 2

The second experiment brought the aiming requirements closer to those in a VMR task by adding in an autocorrection module to the cannon (Figures 2A and 2B). The autocorrection module automatically and covertly corrected for task error in a monotonic fashion (using an SSM; STAR Methods). Participants learned to compensate for all feedback rotations. The initial rapid increase in aiming was followed by a slow decrease. Combined with the autocorrection module, this caused the cannonball to continue going toward the target region (Figures 2C and 2D).

Experiment 2a used a single target position. Participants were exposed to either the ±15°, ±30°, and ±45° or the ±45°, ±60°, and ±90° rotations. During the baseline phase, there was no bias in aiming for any rotation size ($p > 0.441$). The cannonball angle (aim + autocorrection) during the final five trials was not different from the ideal angle in any rotation group ($p = 1$). After the first washout trial, aiming under all rotations was biased in the direction opposite the previously ideal angle ($p < 0.001$). This effect was because of the persisting effects of the autocorrection module, whose learned bias now offset the cannonball angle in the opposing direction ($p < 0.001$).

Experiment 2b used the target layout from 1b, and each participant was exposed to one of ±15°, ±30°, ±45°, ±60°, or ±90° rotations in a task comprising a single block. During the baseline phase, there was no aiming bias for any group ($p > 0.356$). In the middle five cycles of the task, a *t* test versus the ideal angle revealed that participants already fully compensated for the 15° rotation (95% CI, 14.07°, 15.26°, $t(57) = 1.122$, $p = 1$) but not in any other group ($p < 0.02$). The mean (±SD) number of cycles until peak aim was 8.24 (±10.94), 8.12 (±8.56), 9.98 (±10.7), 12.65 (±10.57), and 15.69 (±11.97) for the 15°, 30°, 45°, 60°, and 90°, respectively, such that participants tended to reach their peak aims later in the experiment under larger rotations (Figure 2F, rightmost). During their cycle of peak aim, participants aimed 15.54° (±6.85°), 30.95° (±4.39°), 46.18° (±12.14°), 65.94° (±14.3°), and 94.92° (±7.39°), in their respective groups. By the final rotation cycle, aiming reduced to 3.24° (±3.43°), 18.18° (±4.58°), 27.94° (±8.73°), 47.95° (±11.91°), and 73.68° (±18.41°), respectively.

Direct comparison with VMR task

To explore whether aiming behavior is consistent with that observed in VMR tasks, we now directly compare key properties of learning between experiment 2b and the eight-target experiment of Bond and Taylor (Figures 2E and 2F). We find a general similarity between the two datasets, indicating that insights discovered using Cannon Game may generalize to VMR tasks.

A 2 (experimental paradigm; explicit versus reaching) × 5 (rotation size) × 3 (learning phase) ANOVA on aim angles found a main effect of experimental paradigm ($F(1, 933) = 20.111$,

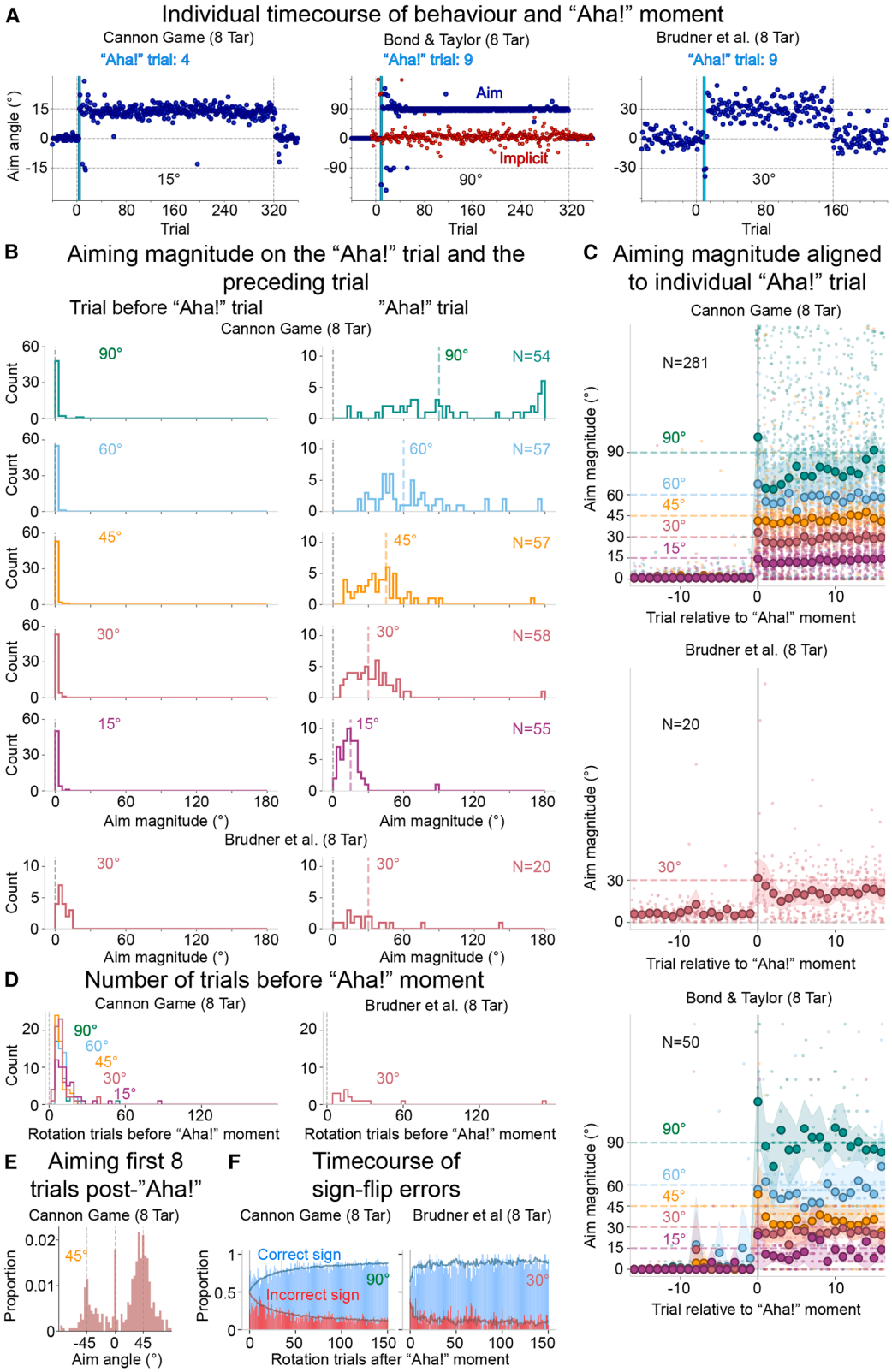
(E) Performance during the rotation phase in experiment 2b (blue) compared with data from Bond and Taylor (2015; orange). Autocorrection model (SSM) parameters were fit to implicit learning in Bond and Taylor (STAR Methods). Autocorrection/implicit learning and cannonball flight/hand angle are included for transparency but de-emphasized as the aiming comparison is primary.

(F) Comparison of summary statistics for each rotation size. In order, from left to right: each participant's greatest cycle-averaged aiming value; each participant's aiming value on the final rotation cycle; each participant's final aim subtracted from their peak aim; the cycle number at which each participant's peak aim was observed.

For (C)–(E), points show group means, and error bands show 95% CIs. For (F), the boxplots show the quartiles of the data, the whiskers bars show 1.5× the interquartile range, and the points show individual data.

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

$p < 0.001$, $\eta_p^2 = 0.021$, 95% CI [0.006, 0.044]), which varied depending on rotation size ($F(4, 933) = 8.434$, $p < 0.001$, $\eta_p^2 = 0.035$, 95% CI [0.017, 0.065]). We found no interaction effect of learning phase × paradigm ($F(2, 933) = 1.165$, $p = 0.312$, $\eta_p^2 < 0.001$, 95% CI [0.000, 0.013]) or the three-way interaction ($F(8, 933) = 0.745$, $p = 0.651$, $\eta_p^2 = 0.006$, 95% CI [0.004, 0.031]). Post hoc pairwise comparisons (simple effect of paradigm at each rotation size) revealed an effect for the 90° rotation (95% CI, 5.8°, 14.7°, $t_{(23.1)} = 4.76$, $p < 0.001$, Cohen's d = 1.116) but for no other rotation ($p > 0.433$). Between the two experiments, there was no difference in the cycle index of peak aim for any rotation ($p > 0.129$; Figure 2F, right). Within this cycle of peak aiming, there was no difference in aiming for any rotation ($p > 0.085$). In the final rotation cycle, there was also no difference in aiming ($p > 0.363$; Figure 2F, second). There was a greater decrease in aiming (the difference between peak aim and final-cycle aim) in Cannon Game (12.3°) than Bond and Taylor (6.89°) for the 15° rotation (95% CI, 1.86°, 8.95°, $t(66) = 3.05$, $p = 0.017$, Cohen's d = 1.04) but no other ($p > 0.087$; Figure 2F, third).

Elucidating the "Aha!" moment

The group learning curve evolves in a gradual, monotonic fashion until reaching asymptote. However, this gradual, monotonic group curve could result from averaging over individual functions, which may not be gradual and monotonic themselves.[14,15]

To investigate this possibility, we began by finding the hypothesized "Aha!" moment for each participant of every dataset, using the model-free pruned exact linear time (PELT) algorithm (STAR Methods). For example, in Figure 3A, we can see an exemplar from 3 different experimental paradigms. On these plots, the "Aha!" trial demarcates a discontinuity between baseline-like aiming behavior and some newly adopted aiming strategy. Each exemplar exhibits sign-flip errors in the early section of their post-"Aha!" aiming, indicating that the "Aha!" moment did not result in an efficacious strategy en toto.

To accommodate the separate timescales in learning sign and magnitude, we look at the aiming magnitude in isolation. Then, to meaningfully aggregate across participants, we realigned the chronology of their aiming data such that each individual's first post-"Aha!" trial was at trial $t = 0$. We can see in Figure 3B that aiming on trial $t = -1$ was tightly grouped near zero; indeed, aiming magnitude during the pre-"Aha!" trial was not different from the baseline in any group in experiments 1b, 2b, Bond and Taylor, or Brudner et al. ($p > 0.12$).

This baseline perseveration was followed by an abrupt single-trial shift to good magnitude performance. Aim magnitudes on the "Aha!" trial, in all experiments (including Bond and Taylor and Brudner et al.), were similar to the ideal magnitude ($p > 0.16$). Some participants intermittently return to their baseline strategy during the early post-"Aha!" phase (Figure 3E), such that the aiming magnitude on the following eight trials was significantly below the ideal (e.g., Figure 3C). The "Aha!" moment typically occurs within a few trials (Figure 3D) and is followed by frequent sign-flip errors (Figure 3F). A sign-flip error is where the aiming magnitude is roughly correct, but the sign is in the wrong direction.

Is there any learning or exploration before the "Aha!" moment?

So far, the data strongly suggest the presence of an "Aha!" moment that manifests in an abrupt single-trial shift to a new strategy. However, it may be that subtle learning or exploration occurs before this apparent "Aha!" moment. This might indicate that the true learning process is one of trial and error. To investigate this, we quantitatively test for any signs of learning or exploration before the "Aha!" moment.

We first tested if there was any difference in aiming variability between the pre-"Aha!" and baseline phases (Figure 4A). To avoid any sample-size bias effects, we used the same number of trials from each phase. If exploration were present, we would expect an increase in aiming variability in the pre-"Aha!" phase. We found no difference in any dataset ($p > 0.075$) apart from Brudner, where a significant decrease was observed (baseline, 8.39° [95% CI, 6.5°, 11.88°]; pre-"Aha!", 5.14° [95% CI, 4.03°, 6.29°]; $t(19) = 2.24$, $p = 0.038$), likely because two participants showed abnormally high variability during baseline (Figure 4A) or because participants continued to improve their baseline-strategy execution during the pre-"Aha!" rotation phase.

We submitted these pre-"Aha!" rotation-phase data to a test of significant error correction rate (i.e., the average trial-wise correction for errors), using one-sample t tests against zero (Figure 4B). If any learning were present, we would expect a positive error correction rate. The correction rate was not different from 0 in any rotation size of any dataset ($p > 0.09$), except that of Brudner et al., which was significantly negative (−0.081, 95% CI, −0.15, −0.05, $t(19) = 2.79$, $p = 0.012$), indicating a small drift in the wrong direction.

Together, these results reinforce our earlier findings that this pre-"Aha!" rotation-phase aiming appears to be no different

Figure 3. An "Aha!" moment precedes the generation of a new strategy

(A) Examples of individual aiming behavior (blue) over time in eight-target vanilla Cannon Game (left), Bond and Taylor (middle), and Brudner et al. (right). Note the change in y-scale. The algorithmically detected "Aha!" trial is highlighted in cyan (see STAR Methods for details on PELT algorithm), and for Bond and Taylor, implicit learning is in red.

(B) All participant aim magnitudes for the trial immediately preceding their respective "Aha!" moment (left) and on the subsequent trial (right), for vanilla Cannon Game (top) and Brudner et al. (bottom).

(C) De-signed aiming magnitude over time for vanilla Cannon Game (top), Brudner et al. (middle), and Bond and Taylor (bottom), where each individual's data have been realigned such that trial 0 is the first trial at which a new strategy has been observed for that participant.

(D) Histogram of the number of rotation-phase trials before each participant's respective "Aha!" moment for vanilla Cannon Game (left) and Brudner et al. (right).

(E) Distribution of all participant's aim signed aim angles for the first eight trials following their respective "Aha!" moments (including the "Aha!" trial), for the 45° group of the eight-target vanilla Cannon Game.

(F) Proportion of aims having the correct (blue) and incorrect (red) sign over time, during the rotation phase, for vanilla Cannon Game (left) and Brudner et al. (right). Solid lines are the "Aha!" model aiming sign proportions. These data are realigned to each individual's respective "Aha!" moment and include only the aims that were >15° away from the target. Error bands = 95% CIs.

See also Figure S1.

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

CellPress
OPEN ACCESS

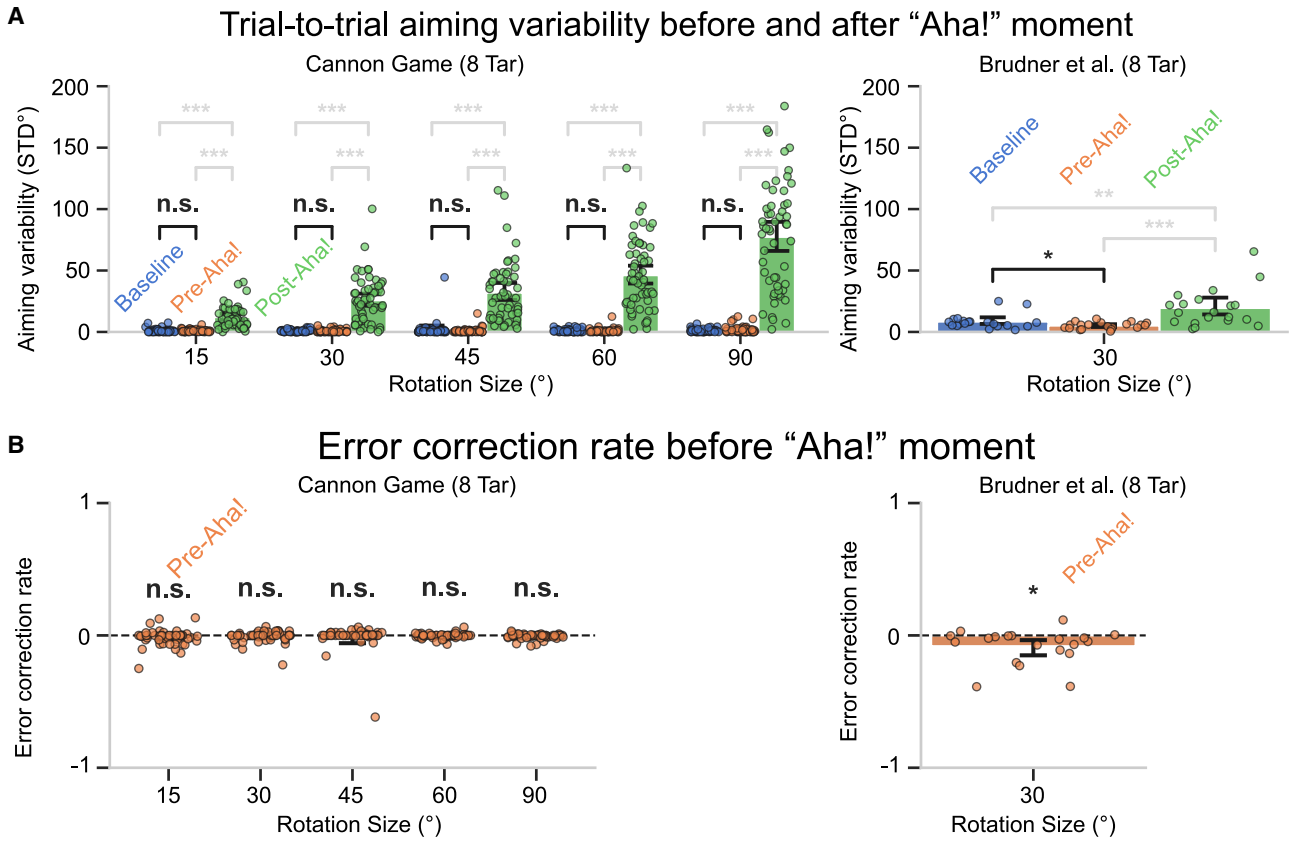


Figure 4. Pre-"Aha!" rotation-phase aiming behavior does not differ from baseline aiming behavior

(A) Trial-to-trial aiming variability during baseline (blue), pre-"Aha!" rotation phase (orange), and post-"Aha!" rotation phase (green) for vanilla Cannon Game (left) and Brudner et al. (right). Significance results of paired *t* tests are shown; the key test result of baseline versus pre-"Aha!" rotation-phase aiming variability is in black.

(B) Error correction rate during pre-"Aha!" rotation-phase for vanilla Cannon Game (left) and Brudner et al. (right). Significance results of one-sample *t* tests are shown. These analyses used matched-window sizes, such that if the participant has n pre-"Aha!" rotation-phase trials, then only the last n baseline and first in post-"Aha!" trials are used. Only participants with at least two pre-"Aha!" rotation trials were included in the variability and error correction analyses.

from baseline aiming and exhibits no signs of exploration or error correction.

Model-based analysis

We now employed model-based analyses to more directly test our "Aha!" hypothesis against other explanations of strategic behavior. We used three models in this analysis: for gradual, error-based learning, we used a single-process SSM (which yields a better BIC than a dual-process SSM; Figure S2); for trial-and-error learning, we introduce a novel variant of the Q-learning algorithm, which learns separate Q-functions for sign and magnitude; and to model the "Aha!" moment, we embed our Q-learning algorithm within a two-state hidden Markov model (HMM). This "Aha!" model must infer that the prediction error is relevant before it learns to compensate, which typically results in a single-trial learning explosion after some arbitrary period of baseline perseverance (Figure S3).

We fit all three models to the first block of each individual's aiming data in all experiments (see STAR Methods for details). Parameter fits for all models can be seen in Figure S3. Of the 777 participants across all datasets, 752 participants were best fit by the "Aha!" model according to BIC scores (Figure 5E; we also fit the models to Ding et al.,[46] and the "Aha!" model yielded the lowest BIC for 549 out of 560 participants; Figure S1G). As can be seen in Figure 5A, the "Aha!" model replicates the abrupt shift from baseline aiming to correctly adjusting its compensatory magnitude for the noisily estimated feedback rotation (see also Figures S1A and S1B). The "Aha!" model also reproduces the pattern of sign flips and intermittent recurrence of the baseline strategy, while the SSM fails to display any of these behaviors (Figures 5B and 5C).

To quantify the fidelity of each model to the human data, we submit the mean rotation-phase predictive policy of each model to a Watson's U^2 test against the respective human distribution. The U^2 metric is a measure of similarity between two cumulative distribution functions (CDFs) on the circle (lower is better). The Bond and Taylor dataset was excluded from this analysis as the possible aim locations were in increments of 5.625°, leading to aberrant behavior. Watson's U^2 favored the "Aha!" model across all rotation sizes in all experiments, including Ding et al. ($t > 4.03$, $p < 0.001$, Cohen's $d > 0.65$).

These findings, combined with the visualizations in Figures 5 and S1, further evidence the presence of an "Aha!" moment. Moreover, they indicate that our "Aha!" model accurately

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

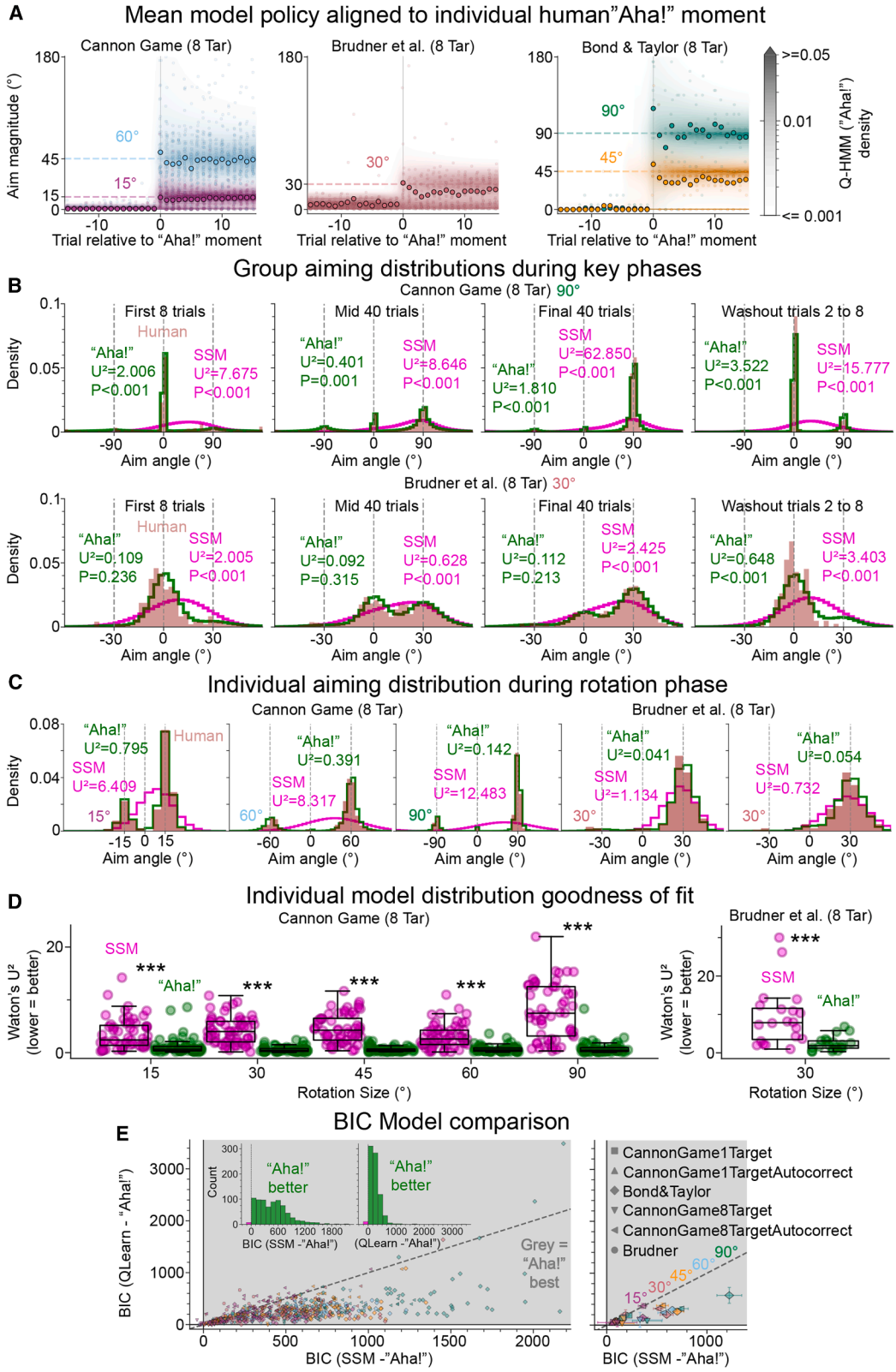


Figure 5. Human aiming behavior is more similar to the "Aha!" model than to gradual error-based learning or Q-learning

(A) Human aiming magnitude (points) and mean "Aha!" model predictive policy distribution (contours) over time, for vanilla Cannon Game (left), Brudner et al. (middle), and Bond and Taylor (right), using data realigned such that trial 0 is the first trial after each individual human's algorithmically detected "Aha!" moment.

(legend continued on next page)

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

captures much of human behavior, whereas gradual error minimization or mere trial-and-error learning does not.

DISCUSSION

Strategic behavior in VMR tasks is commonly characterized as either a gradual error-minimization process or as a process of trial-and-error learning. We sought to test these hypotheses on isolated strategic behavior. We found that individual learning functions often do not gradually increase toward the ideal aim angle, nor do they typically appear to explore various strategies as predicted by typical trial-and-error learning accounts. Instead, most individuals perseverate in aiming at the target before experiencing an "Aha!" moment and suddenly shifting their aiming magnitude to compensate for a noisily estimated feedback perturbation. The correct sign often takes additional time to be learned. Some people do adopt a strategy that is entirely unlike the optimal strategy, but they still exhibit the same "Aha!" moment before its adoption.

We re-examined data from a delayed-feedback VMR task, which had no aim reports or prompts,[45] and uncovered the same hallmarks of an "Aha!" moment. We find further support for the "Aha!" hypothesis in a recent dataset,[46] although the "Aha!" moment was expedited (and likely corrupted) by an explicit instruction to explore aiming strategies. Model-based analyses corroborate the presence of an "Aha!" moment.

Strategic behavior without motor contamination

We isolated strategic behavior, in a cannon-shooting game, removed from motor contaminants. In experiments 1, we established that behavior was generally consistent with that in VMR literature; we replicated the gradual, monotonic improvement in group-level performance followed by asymptote (Figure 1B).

To bring the aiming requirements of our task closer to those in a VMR task, we added an autocorrection module into the cannon. This covertly learned to gradually reduce the discrepancy between the aim and feedback in a similar manner to implicit adaptation (Figures 2A and 2B; see STAR Methods for details). This autocorrection used plan-based generalization of implicit adaptation,[47–49] although simulations found no major differences if target-based generalization was used instead.

We then characterized the degree of similarity between Cannon Game data and VMR data where performance is broken down into aim reports and implicit adaptation.[1] Aiming behavior in experiment 2b did not differ from Bond and Taylor during the early-, mid-, or late-rotation phases under any rotation size except 90°. This is consistent with strategic behavior correcting for errors left by the sensorimotor system.[8] We next defined three essential properties to characterize an individual's learning curve: the cycle of peak aim, the peak-cycle-averaged aim, and the decrease in aim from the peak to final cycle. These were generally similar across experiments (Figure 2F), prompting an investigation into individual learning.

The individual time course of aiming

In contrast to the canonical learning curve, the typical individual tends to perseverate in aiming at the target until abruptly shifting his or her aim to approximate compensatory magnitude of the noisily estimated feedback rotation (e.g., Figure 3B). These disparate patterns of behavior are partly reconciled by the fact that taking the mean over variably delayed step-functions yields a smooth curve.[14,15] Another contributing factor to this artefactual group-level curve is that participants typically take additional time after their "Aha!" moment to reliably use the correct sign and intermittently revert to baseline behavior (Figures 3E and 3F).

Alternative accounts of strategic behavior

Strategic behavior has traditionally been modeled as a gradual error-minimization process via an SSM.[3,8,11–13] As can be seen in Figures 3A and S4A, there is a clear discontinuity in aiming strategy and no sign of prior gradual error minimization (Figure 4B). Aligning individual time courses to their "Aha!" moments demonstrate this effect at the group level (Figures 3C and S4B). While there may still be error-based fine-tuning of the strategy, strategy adoption seems discrete.

Another competing account is that strategic re-aiming in VMR tasks reflects a process of exploration until an efficacious action is found and can be exploited.[50] This RL account predicts exploration of competing strategies before settling on the correct one. However, as can be seen in Figures 4, S1C, and S1D, there is no evidence of an increase in trial-to-trial aiming variability, or error correction rate, during the pre-"Aha!" rotation phase.

Another recent account proposes that strategic behavior reflects a process of hypothesis testing.[10,46] This entails the generation of hypotheses about the nature of the environment and testing their efficacy. As with the RL account, we do not view this as being at odds with our "Aha!" moment hypothesis; rather, they are focused on different aspects of learning. Central to our

(B) Group-level distributions of signed human aiming (red), "Aha!" model predictive density (green) and SSM predictive density (magenta), during the first eight rotation trials (left), the next 40 trials (middle left), the final 40 rotation trials (middle right), and the second to eighth washout trials (right), for vanilla Cannon Game (top) and Brudner et al. (bottom). Watson's U^2 test statistic is printed on each panel, for each model versus the human aims, which measures deviance from homogeneity between two circular distributions (lower = better).

(C) Individual human signed aiming distributions (red), "Aha!" model predictive density (green), and SSM predictive density (magenta), for the entire rotation phase, for vanilla Cannon Game (left) and Brudner et al. (right).

(D) Watson's U^2 test statistic values for all individuals in vanilla Cannon Game (left) and Brudner et al. (right), for the "Aha!" model (green) and the SSM (magenta). Significance results for paired t tests between SSM and "Aha!" models' U^2 are shown (black stars).

(E) Relative BIC scores for the three fitted models for individuals (left) and group means (right), for all six datasets (vanilla Cannon Game one-target [square] and eight-target [diamond]; autocorrecting Cannon Game one-target [up triangle] and eight-target [left triangle]; Brudner et al. [circle]; Bond and Taylor [down triangle]). The gray quadrant denotes a better fit for the "Aha!" model than for the SSM or the Q-learning model. The identity line indicates the threshold for Q-learning versus SSM (below the identity line denotes that Q-learning fit better than the SSM). Inset: number of participants within each range of relative BIC scores for the SSM versus "Aha!" (left) and Q-learning versus "Aha!" (right); green = "Aha!" model better, and magenta = "Aha!" model worse. Error bars are 95% CIs.

See also Figures S2–S5.

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

hypothesis is that the feedback error is considered uninformative to the agent's goals until some moment of insight occurs and the participant's internal representation of the task changes. This appears to manifest in a strategic solution, which is correct in at least the magnitude dimension (Figures 3B and 3C). The hypothesis-testing account is focused on what happens after the "Aha!" moment.

A moderate proportion of participants from Ding et al.[46] do appear, however, to adopt an incorrect strategy at their "Aha!" moment (e.g., aiming 180° away; Figure S1F). Importantly, these participants were explicitly instructed to explore new aiming strategies to compensate for a change in action-feedback mapping during the rotation phase. This instruction appears to have expedited and corrupted the "Aha!" moment by encouraging premature strategy exploration (for example, 145 participants changed their strategy on the first rotation trial before observing any rotated feedback; Figure S1E).

Overall, our analyses indicate that the "Aha!"-moment account explains much of the strategic behavior in VMR tasks, leaving a relatively smaller portion for RL or hypothesis testing.

The generative "Aha!" model

We developed a generative model that combines hierarchical off-policy Q-learning with a two-state HMM: in one state, the prediction error is considered irrelevant to the agent's goals, and in the other state, the prediction error is considered relevant. In this way, the model can perseverate in baseline behavior due to a strong prior that the prediction error is irrelevant. The dense rewards and off-policy nature of learning allow the model to shift from baseline aiming on trial t to the ideal aim on trial $t + 1$ (Figure S3). The off-policy updates allow the model to counterfactually generalize prediction error across the workspace, akin to counterfactual reasoning observed in humans.[51,52] It learns separate Q-functions for sign and magnitude, recapitulating the pattern of sign-flip errors in the post-"Aha!" rotation phase (Figures 5B, 5C, S4H, and S4I).

State inference within the model is subject to uncertainty over the prediction error. This context uncertainty increases with rotation size (Figure S4A), which aligns well with the fact that rotation size had no effect on the delay before an "Aha!" moment, despite the larger rotations providing a stronger error signal (Figure 3D). We know from Zhang et al.[53] that the standard deviation of visual uncertainty increases in a roughly linear fashion (see also Levi et al.[54] and Klein and Levi[55]). Indeed, we fit a linear function to context uncertainty and found a slope like that found by Zhang et al. (Figure S5).

We also see an exponential decay in inverse temperature as rotation size increases (Figure S4A). In RL, the inverse temperature defines the trade-off between exploration and exploitation. For our model, this means that larger rotation sizes, which elicit greater context uncertainty, see greater behavioral variability. However, it may be the case that the increased variability reflects the difficulty in accurately estimating the prediction error or reflects bet-hedging against the uncertainty therein, rather than exploration over varied strategies.

When present reward can augment future reward, hedging against uncertainty maximizes the long-term reward in a varying environment.[56,57] For example, it is an optimal strategy for maximizing long-term returns under variable financial market conditions[58,59] or for maximizing population fitness in dynamic adverse environments.[60,61] Many natural human tasks do involve rewards, which can augment future rewards. For example, any task that results in food, water, or social status thus provides sustenance to better perform subsequent tasks. Therefore, it may be that the increase in exploration associated with an increase in context uncertainty reflects a form of hedging to maximize lifelong geometric-mean reward (see also Kaufmann et al.[62] for proof of Thompson sampling optimality for cumulative regret minimization). Work by Feng et al.[63] is consistent with this hypothesis and finds that exploration is driven by the signal-to-noise ratio of reward information. Thus, we do not see good reason to believe that the variable compensatory magnitude reflects the exploration of strategies.

The "Aha!" model isn't intended to be a complete mechanistic account of the "Aha!" moment. Nevertheless, the model can generally capture the relevant phenomena covered in this work. While RL is essential to this model, the RL component relies on a counterfactual reward to direct its behavior, so we do not consider the initial strategy generation as being part of a trial-and-error learning process.

Out of 1,337 participants in our modeling analysis, including data collected in VMR reaching tasks with no explicit aim reports, 1,301 were best fit by our "Aha!" model (Figures 5E and S4G). Furthermore, only the "Aha!" model was able to reproduce the baseline perseveration along with the multimodal distribution of aims (e.g., Figures 5B, 5C, S4H, and S4I). Taken together, we see this as overwhelming support for our "Aha!" model being superior to extant alternatives.

The "Aha!" moment in the motor context

It is possible that our task yields the same behavioral patterns as in standard VMR tasks but through distinct strategic processes. To investigate this concern, we carried out the same model-fitting analysis on preexisting reaching data from a task without aim reports and without any prompt that an aiming solution might exist or that one should consider aiming.[45] Instead, the task presented feedback after a delay period of 5 s, which is thought to inhibit implicit adaptation and emphasize strategic contributions.[64] All 20 participants were best fit by the "Aha!" model, and their behavior exhibited the same perseveration and abrupt shift (Figure 3C). Further, we see no reason, *a priori*, to think that strategic processing would be qualitatively—not just quantitatively—affected by this delay.

We also fit the models to recent data from Ding et al.,[46] where participants reached as in typical VMR tasks, although with the explicit instruction to explore new strategies. The "Aha!" model still better explained the behavior of 549 out of 560 participants, and it still matched the general behavioral patterns and distributions (Figure S1).

By incorporating our insights into strategic behavior back into the motor adaptation context, we can attack the problem of implicit learning in new ways. For example, by using our generative "Aha!" model (or a successor) to explain away the portion of behavior attributable to the strategic system, we may be able to investigate implicit adaptation in more ecologically relevant contexts than using error clamps or delayed feedback.

Conclusions

Together, these findings indicate that strategic behavior, both in the Cannon Game and in VMR reaching tasks, is not a gradual error-minimization process or mere trial-and-error learning. Instead, there is perseveration of baseline behavior, followed by an "Aha!" moment that marks a dramatic single-trial shift in behavior. This typically manifests in a correct compensatory magnitude for a noisily estimated rotation, but reliably correct sign behavior takes longer to learn. This "Aha!" moment appears to eat much of the lunch of the trial-and-error accounts. This is important not only for understanding how strategies are acquired and evolve over time but also for attempting to understand implicit adaptation and motor behavior more generally.

Future work should try to better understand the processes that lead to this sudden shift in behavior. Is there a fundamental change in the representation of the task, or does the preceding perseveration reflect the time taken to infer the presence of a new context? Answering this would also expedite the motor rehabilitation process by helping patients to learn strategic behaviors more quickly and effectively.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Max Townsend (max.o.b.townsend@gmail.com).

Materials availability

This study did not generate any new, unique reagents.

Data and code availability

- All data have been deposited at GitHub and are publicly available at https://github.com/Max-Townsend/aha_precedes_strategy_VMR as of the date of publication.
- All original code has been deposited at GitHub and is publicly available at https://github.com/Max-Townsend/aha_precedes_strategy_VMR as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Jordan Taylor for providing the raw data used to compare Gannon Game with VMR data.[1,45] We also thank Wei Ding, Anjuli Niyogi, and Jonathan Tsay for providing the raw data from their VMR task.[46] Finally, we thank Rich Ivry for shepherding the Cannon Game idea through its early years. M.T. was supported by an ESRC White Rose Doctoral Training Partnership Scholarship. F.M. was supported by the UK Research and Innovation Biotechnology and Biological Sciences Research Council (BB/X008428/1). F.M. and M.M.-W. were supported by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (NIHR203331).

AUTHOR CONTRIBUTIONS

Conceptualization, M.T. and J.R.M.; methodology, M.T., F.M., and J.R.M.; software, M.T.; validation, M.W.; formal analysis, M.T.; investigation, M.T.; modeling, M.T.; writing – original draft, M.T.; writing – review and editing, M.T., M.W., C.C., M.M.W., F.M., and J.R.M.; visualization, M.T.; supervision, M.W., C.C., M.M.W., F.M., and J.R.M.; funding acquisition, M.M.W., F.M., and J.R.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
- METHOD DETAILS
  - Experimental Apparatus
  - Aiming Task
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data Analysis
  - Data-driven "Aha!" moment detection
  - Computational models

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2026.04.021.

Received: August 6, 2025
Revised: March 5, 2026
Accepted: April 11, 2026

REFERENCES

1. Bond, K.M., and Taylor, J.A. (2015). Flexible explicit but rigid implicit learning in a visuomotor adaptation task. J. Neurophysiol. *113*, 3836–3849. https://doi.org/10.1152/jn.00009.2015.
2. McDougle, S.D., Ivry, R.B., and Taylor, J.A. (2016). Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. Trends Cogn. Sci. *20*, 535–544. https://doi.org/10.1016/j.tics.2016.05.002.
3. McDougle, S.D., Bond, K.M., and Taylor, J.A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. J. Neurosci. *35*, 9568–9579. https://doi.org/10.1523/JNEUROSCI.5061-14.2015.
4. McDougle, S.D., and Taylor, J.A. (2019). Dissociable cognitive strategies for sensorimotor learning. Nat. Commun. *10*, 40. https://doi.org/10.1038/s41467-018-07941-0.
5. Morehead, J.R., Butcher, P.A., and Taylor, J.A. (2011). Does fast learning depend on declarative mechanisms? J. Neurosci. *31*, 5184–5185. https://doi.org/10.1523/JNEUROSCI.0040-11.2011.
6. Taylor, J.A., Krakauer, J.W., and Ivry, R.B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. J. Neurosci. *34*, 3023–3032. https://doi.org/10.1523/JNEUROSCI.3619-13.2014.
7. Taylor, J.A., Klemfuss, N.M., and Ivry, R.B. (2010). An explicit strategy prevails when the cerebellum fails to compute movement errors. Cerebellum *9*, 580–586. https://doi.org/10.1007/s12311-010-0201-x.
8. Taylor, J.A., and Ivry, R.B. (2011). Flexible cognitive strategies during motor learning. PLoS Comput. Biol. *7*, e1001096. https://doi.org/10.1371/journal.pcbi.1001096.
9. Taylor, J.A., and Ivry, R.B. (2012). The role of strategies in motor learning. Ann. N. Y. Acad. Sci. *1251*, 1–12. https://doi.org/10.1111/j.1749-6632.2011.06430.x.
10. Tsay, J.S., Kim, H.E., McDougle, S.D., Taylor, J.A., Haith, A., Avraham, G., Krakauer, J.W., Collins, A.G.E., and Ivry, R.B. (2024). Fundamental processes in sensorimotor learning: Reasoning, refinement, and retrieval. eLife *13*, e91839. https://doi.org/10.7554/eLife.91839.
11. Takiyama, K., Hirashima, M., and Nozaki, D. (2015). Prospective errors determine motor learning. Nat. Commun. *6*, 5925. https://doi.org/10.1038/ncomms6925.

12. Coltman, S.K., Cashaback, J.G.A., and Gribble, P.L. (2019). Both fast and slow learning processes contribute to savings following sensorimotor adaptation. J. Neurophysiol. *121*, 1575–1583. https://doi.org/10.1152/jn.00794.2018.

13. Zhang, X., Wu, W., Wei, K. Perceptual prediction error supports implicit process in motor learning. PLoS Comput. Biol. 22:e1014196. https://doi.org/10.1371/journal.pcbi.1014196.

14. Bahrick, H.P., Fitts, P.M., and Briggs, G.E. (1957). Learning curves; facts or artifacts? Psychol. Bull. *54*, 255–268. https://doi.org/10.1037/h0040313.

15. Gallistel, C.R., Fairhurst, S., and Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. Proc. Natl. Acad. Sci. USA *101*, 13124–13131. https://doi.org/10.1073/pnas.0404965101.

16. Collins, A., and Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. PLoS Biol. *10*, e1001293. https://doi.org/10.1371/journal.pbio.1001293.

17. Darshan, R., Leblois, A., and Hansel, D. (2014). Interference and shaping in sensorimotor adaptations with rewards. PLoS Comput. Biol. *10*, e1003377. https://doi.org/10.1371/journal.pcbi.1003377.

18. Donoso, M., Collins, A.G.E., and Koechlin, E. (2014). Human cognition. Foundations of human reasoning in the prefrontal cortex. Science *344*, 1481–1486. https://doi.org/10.1126/science.1252254.

19. Griffiths, T.L., Chater, N., Kemp, C., Perfors, A., and Tenenbaum, J.B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. Trends Cogn. Sci. *14*, 357–364. https://doi.org/10.1016/j.tics.2010.05.004.

20. Griffiths, T.L., Callaway, F., Chang, M.B., Grant, E., Krueger, P.M., and Lieder, F. (2019). Doing more with less: meta-reasoning and meta-learning in humans and machines. Curr. Opin. Behav. Sci. *29*, 24–30. https://doi.org/10.1016/j.cobeha.2019.01.005.

21. Piantadosi, S.T., Tenenbaum, J.B., and Goodman, N.D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. Psychol. Rev. *123*, 392–424. https://doi.org/10.1037/a0039980.

22. Nikooyan, A.A., and Ahmed, A.A. (2015). Reward feedback accelerates motor learning. J. Neurophysiol. *113*, 633–646. https://doi.org/10.1152/jn.00032.2014.

23. Izawa, J., and Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. PLoS Comput. Biol. *7*, e1002012. https://doi.org/10.1371/journal.pcbi.1002012.

24. Sugiyama, T., Schweighofer, N., and Izawa, J. (2023). Reinforcement learning establishes a minimal metacognitive process to monitor and control motor learning performance. Nat. Commun. *14*, 3988. https://doi.org/10.1038/s41467-023-39536-9.

25. Carpenter, W. (2020). The Aha! moment: the science behind creative insights. In Toward Super-Creativity - Improving Creativity in Humans, Machines, and Human-Machine Collaborations, S.M. Manuel Brito, ed. (IntechOpen). https://doi.org/10.5772/intechopen.84973.

26. Jones, G. (2003). Testing two cognitive theories of insight. J. Exp. Psychol. Learn. Mem. Cogn. *29*, 1017–1027. https://doi.org/10.1037/0278-7393.29.5.1017.

27. Knoblich, G., Ohlsson, S., Haider, H., and Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. J. Exp. Psychol. Learn. Mem. Cogn. *25*, 1534–1555. https://doi.org/10.1037/0278-7393.25.6.1534.

28. Knoblich, G., Ohlsson, S., and Raney, G.E. (2001). An eye movement study of insight problem solving. Mem. Cognit. *29*, 1000–1009. https://doi.org/10.3758/BF03195762.

29. Kounios, J., and Beeman, M. (2009). The *Aha*! moment: the cognitive neuroscience of insight. Curr. Dir. Psychol. Sci. *18*, 210–216. https://doi.org/10.1111/j.1467-8721.2009.01638.x.

30. Sprugnoli, G., Rossi, S., Emmendorfer, A., Rossi, A., Liew, S.-L., Tatti, E., Di Lorenzo, G., Pascual-Leone, A., and Santarnecchi, E. (2017). Neural correlates of Eureka moment. Intelligence *62*, 99–118. https://doi.org/10.1016/j.intell.2017.03.004.

31. Jung-Beeman, M., Bowden, E.M., Haberman, J., Frymiare, J.L., Arambel-Liu, S., Greenblatt, R., Reber, P.J., and Kounios, J. (2004). Neural activity when people solve verbal problems with insight. PLoS Biol. *2*, E97. https://doi.org/10.1371/journal.pbio.0020097.

32. Novick, L.R., and Sherman, S.J. (2003). On the nature of insight solutions: evidence from skill differences in anagram solution. Q. J. Exp. Psychol. A *56*, 351–382. https://doi.org/10.1080/02724980244000288.

33. Salvi, C., Bricolo, E., Kounios, J., Bowden, E., and Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. Think. Reason. *22*, 443–460. https://doi.org/10.1080/13546783.2016.1141798.

34. MacGregor, J.N., Ormerod, T.C., and Chronicle, E.P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. J. Exp. Psychol. Learn. Mem. Cogn. *27*, 176–201. https://doi.org/10.1037/0278-7393.27.1.176.

35. Sawyer, R.K. (2012). Explaining Creativity: The Science of Human Innovation, Second Edition (Oxford University Press).

36. Vartanian, O., and Goel, V. (2007). Neural correlates of creative cognition. In Evolutionary and Neurocognitive Approaches to Aesthetics, Creativity and the Arts, C. Martindale, P. Locher, V.M. Petrov, and A. Berleant, eds. (Routledge), pp. 195–207.

37. Manley, H., Dayan, P., and Diedrichsen, J. (2014). When money is not enough: awareness, success, and variability in motor learning. PLoS One *9*, e86580. https://doi.org/10.1371/journal.pone.0086580.

38. Sergent, C., and Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. Psychol. Sci. *15*, 720–728. https://doi.org/10.1111/j.0956-7976.2004.00748.x.

39. Trope, Y., and Liberman, N. (2010). Construal-level theory of psychological distance. Psychol. Rev. *117*, 440–463. https://doi.org/10.1037/a0018963.

40. Hegele, M., and Heuer, H. (2010). Implicit and explicit components of dual adaptation to visuomotor rotations. Conscious. Cogn. *19*, 906–917. https://doi.org/10.1016/j.concog.2010.05.005.

41. Langsdorf, L., Maresch, J., Hegele, M., McDougle, S.D., and Schween, R. (2021). Prolonged response time helps eliminate residual errors in visuomotor adaptation. Psychon. Bull. Rev. *28*, 834–844. https://doi.org/10.3758/s13423-020-01865-x.

42. Maresch, J., Mudrik, L., and Donchin, O. (2021). Measures of explicit and implicit in motor learning: what we know and what we don't. Neurosci. Biobehav. Rev. *128*, 558–568. https://doi.org/10.1016/j.neubiorev.2021.06.037.

43. Maresch, J., Werner, S., and Donchin, O. (2021). Methods matter: Your measures of explicit and implicit processes in visuomotor adaptation affect your results. Eur. J. Neurosci. *53*, 504–518. https://doi.org/10.1111/ejn.14945.

44. Morehead, J.R., Taylor, J.A., Parvin, D.E., and Ivry, R.B. (2017). Characteristics of implicit sensorimotor adaptation revealed by task-irrelevant clamped feedback. J. Cogn. Neurosci. *29*, 1061–1074. https://doi.org/10.1162/jocn_a_01108.

45. Brudner, S.N., Kethidi, N., Graeupner, D., Ivry, R.B., and Taylor, J.A. (2016). Delayed feedback during sensorimotor learning selectively disrupts adaptation but not strategy use. J. Neurophysiol. *115*, 1499–1511. https://doi.org/10.1152/jn.00066.2015.

46. Ding, W., Niyogi, A., and Tsay, J.S. (2025). Hypothesis testing governs an efficiency-flexibility trade-off in strategic motor learning. Preprint at bioRxiv. https://doi.org/10.64898/2025.11.29.691289.

47. McDougle, S.D., Bond, K.M., and Taylor, J.A. (2017). Implications of plan-based generalization in sensorimotor adaptation. J. Neurophysiol. *118*, 383–393. https://doi.org/10.1152/jn.00974.2016.

48. Day, K.A., Roemmich, R.T., Taylor, J.A., and Bastian, A.J. (2016). Visuomotor learning generalizes around the intended movement. eNeuro *3*. ENEURO.0005-16.2016. https://doi.org/10.1523/ENEURO.0005-16.2016.

49. Schween, R., Taylor, J.A., and Hegele, M. (2018). Plan-based generalization shapes local implicit adaptation to opposing visuomotor

transformations. J. Neurophysiol. *120*, 2775–2787. https://doi.org/10.1152/jn.00451.2018.

50. Cashaback, J.G.A., McGregor, H.R., Mohatarem, A., and Gribble, P.L. (2017). Dissociating error-based and reinforcement-based loss functions during sensorimotor learning. PLoS Comput. Biol. *13*, e1005623. https://doi.org/10.1371/journal.pcbi.1005623.

51. Byrne, R.M.J. (2002). Mental models and counterfactual thoughts about what might have been. Trends Cogn. Sci. *6*, 426–431. https://doi.org/10.1016/S1364-6613(02)01974-5.

52. Van Hoeck, N., Watson, P.D., and Barbey, A.K. (2015). Cognitive neuroscience of human counterfactual reasoning. Front. Hum. Neurosci. *9*, 420. https://doi.org/10.3389/fnhum.2015.00420.

53. Zhang, Z., Wang, H., Zhang, T., Nie, Z., and Wei, K. (2024). Perceptual error based on Bayesian cue combination drives implicit motor adaptation. eLife *13*, RP94608. https://doi.org/10.7554/eLife.94608.

54. Levi, D.M., Klein, S.A., and Yap, Y.L. (1987). Positional uncertainty in peripheral and amblyopic vision. Vis. Res. *27*, 581–597. https://doi.org/10.1016/0042-6989(87)90044-7.

55. Klein, S.A., and Levi, D.M. (1987). Position sense of the peripheral retina. J. Opt. Soc. Am. A *4*, 1543–1553. https://doi.org/10.1364/JOSAA.4.001543.

56. Philippi, T., and Seger, J. (1989). Hedging one's evolutionary bets, revisited. Trends Ecol. Evol. *4*, 41–44. https://doi.org/10.1016/0169-5347(89)90138-9.

57. Starrfelt, J., and Kokko, H. (2012). Bet-hedging—a triple trade-off between means, variances and correlations. Biol. Rev. Camb. Philos. Soc. *87*, 742–755. https://doi.org/10.1111/j.1469-185X.2012.00225.x.

58. Markowitz, H. (1952). Portfolio selection. J. Finance *7*, 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

59. Jorion, P. (2009). Risk management lessons from the credit crisis. Euro. Fin. Manag. *15*, 923–933. https://doi.org/10.1111/j.1468-036X.2009.00507.x.

60. Hopper, K.R. (1999). Risk-spreading and bet-hedging in insect population biology. Annu. Rev. Entomol. *44*, 535–560. https://doi.org/10.1146/annurev.ento.44.1.535.

61. Olofsson, H., Ripa, J., and Jonzén, N. (2009). Bet-hedging as an evolutionary game: the trade-off between egg size and number. Proc. Biol. Sci. *276*, 2963–2969. https://doi.org/10.1098/rspb.2009.0500.

62. Kaufmann, E., Korda, N., and Munos, R. (2012). Thompson sampling: An asymptotically optimal finite-time analysis. In International Conference on Algorithmic Learning Theory, pp. 199–213.

63. Feng, S.F., Wang, S., Zarnescu, S., and Wilson, R.C. (2021). The dynamics of explore–exploit decisions reveal a signal-to-noise mechanism for random exploration. Sci. Rep. *11*, 3077. https://doi.org/10.1038/s41598-021-82530-8.

64. Schween, R., and Hegele, M. (2017). Feedback delay attenuates implicit but facilitates explicit adjustments to a visuomotor rotation. Neurobiol. Learn. Mem. *140*, 124–133. https://doi.org/10.1016/j.nlm.2017.02.015.

65. Krakauer, J.W., Pine, Z.M., Ghilardi, M.F., and Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. J. Neurosci. *20*, 8916–8924. https://doi.org/10.1523/JNEUROSCI.20-23-08916.2000.

66. Mattar, A.A.G., and Ostry, D.J. (2007). Modifiability of generalization in dynamics learning. J. Neurophysiol. *98*, 3321–3329. https://doi.org/10.1152/jn.00576.2007.

67. Taylor, J.A., and Ivry, R.B. (2013). Context-dependent generalization. Front. Hum. Neurosci. *7*, 171. https://doi.org/10.3389/fnhum.2013.00171.

68. Killick, R., Fearnhead, P., and Eckley, I.A. (2012). Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. *107*, 1590–1598. https://doi.org/10.1080/01621459.2012.737745.

69. Truong, C., Oudre, L., and Vayatis, N. (2020). Selective review of offline change point detection methods. Signal Process. *167*, 107299. https://doi.org/10.1016/j.sigpro.2019.107299.

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Data and code for analysis	This paper	https://github.com/Max-Townsend/aha_precedes_strategy_VMR
Software and algorithms		
Unity game engine	Unity Technologies	https://unity.com/
Python 3.11	Python Software Foundation	https://www.python.org/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

A total of 761 humans (316 males, 440 females, 5 other genders, age = 34.1 ± 11.8 years) from the United Kingdom and the United States were recruited via the Prolific Academic recruitment platform. Ethical approval was obtained from the University of Leeds school of Psychology Research Ethics Committee (ref PSYC-81). All participants gave informed consent prior to starting the study.

METHOD DETAILS

Experimental Apparatus

Participants aimed a cannon and shot targets using keyboard keys. All participants completed the task remotely on their own desktop or laptop computers, with no further constraints on the hardware they used. The experimental task was constructed using the Unity game engine and default libraries, with custom scripts written in C# and data saved to a remote database.

Aiming Task

Participants were instructed to rotate the cannon by pressing the 'A' and 'D' keys, and to fire it with the 'Spacebar' key (Figure 1A). They were told to hit as many targets as possible and were not encouraged to act quickly. Tapping the rotation keys rotated the cannon by 1° increments and holding down these keys rotated the cannon at the speed of 125°sec-1. An aiming reticule indicated the aimed location, whose radial distance was held constant (equal to the radial distance of the target). Upon firing, the cannonball flight lasted 300 ms and would terminate at the same radial distance as the target. The target had a radius of 3°, and the cannonball had a radius of 0.5°. Upon being hit, the target exploded and was accompanied by a green 'Hit!' text lasting 500 ms; if missed, a 'buzzing' sound accompanied red 'Miss!' text lasting 500 ms. The Cannon Game task would only play in full-screen mode. At any point, if the participant left full-screen mode, the task was paused and the participant was presented with a pause screen asking them to enter full-screen when they were ready to resume.

Participants were told 'the cannonball may not always go exactly where you are aiming'. The direction of the shot was subject to noise of similar magnitude to motor noise in reaching tasks (Figure 1B), such that the cannonball endpoint feedback was

$$EndPoint^{\circ} = aim^{\circ} + Rotation^{\circ} + \epsilon^{\circ}, \epsilon^{\circ} \sim N(0,3)^{\circ} \quad \text{(Equation 1)}$$

where the integer *aim* is the angle aimed, ϵ is gaussian noise, and *Rotation* is the feedback rotation applied to the endpoint feedback. At the start of each trial, the cannon direction was sampled from a uniform random distribution spanning 60° either side of, but at least 5° away from, the angle equal to $-Rotation^{\circ}$ (i.e., the ideal aim angle).

Experiment 1

The first experiment sought to explore how participants updated their aiming behavior in a simple task which requires the participant to aim at targets, much like in a visuomotor rotation (VMR) task.

In experiment 1a (n = 98), the target was presented in the same position (12 o'clock) on every trial. The experiment comprised six blocks, where each block included 30 rotation trials with a rotation applied to the direction of feedback. The rotation phase was sandwiched between 15 baseline and 15 washout trials, each with veridical feedback (Figure 1A). Within each block, the rotation phase presented a single feedback rotation for all trials, such that one group of participants each experienced every sign/magnitude combination of $\{\pm 15, \pm 30, \pm 45\}^{\circ}$, and the other group experienced all of $\{\pm 45, \pm 60, \pm 90\}^{\circ}$ by the end of the six experimental blocks. Three participants were excluded for having >20% trials with <100ms aiming time.

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

In experiment 1b (n = 301), the target was presented in one of eight locations, evenly spaced in 45° increments from the target at 12 o'clock. The order of target location was pseudorandomized, where each location occurred once in each eight-trial cycle. There was a single block of 40 cycles of rotated feedback, sandwiched between five baseline and five washout cycles. During the rotation phase, each participant experienced one rotation, sampled from $\{\pm 15, \pm 30, \pm 45, \pm 60, \pm 90\}^\circ$. Twenty participants were excluded for having >20% trials with <100ms aiming time.

Experiment 2

Experiment 2 implemented aiming requirements which were much more like those in VMR tasks. To this end, it included an autocorrection module (i.e., a spatially-generalizing SSM adapted from McDougle et al.[47]) which learns a bias that is added to the feedback direction. This bias gradually minimizes the difference between the aimed location and the endpoint feedback (Figures 2A and 2B). The resultant formula for the endpoint feedback direction is

$$EndPoint^\circ = aim^\circ + Rotation^\circ + \overrightarrow{x_t}(aim)^\circ + \epsilon^\circ, \epsilon^\circ \sim N(0,3)^\circ \quad \text{(Equation 2)}$$

where $\overrightarrow{x_t}(aim)^\circ$ is the output of the autocorrection module for integer angle *aim* on trial t. The state vector $\overrightarrow{x_t}$ is updated after each trial t according to

$$\overrightarrow{x_{t+1}} = A\overrightarrow{x_t} - B \exp\left[(\overrightarrow{x} - \mu_t)^2 \Big/ 2\sigma^2\right] \times e_t \quad \text{(Equation 3)}$$

where $\overrightarrow{x_{t+1}}$ is the state vector on trial t+1, A is the retention factor, B is the learning rate, e_t is the prediction error $(-AimAngle_t)$, $\overrightarrow{x}$ is the ordered vector of all integer angles around the workspace (-179,-178…,180), and μ_t is the centre of gaussian generalization on trial t. The parameter μ_t is set to the aim angle on trial t, such that the generalisation of learning is centred on the aimed angle (Figures 2A and 2B). Parameter σ, which defines the width of generalisation, was set to 30° (selected to be consistent with previous generalization widths[47,65–67]).

For Experiment 2a (n = 61), the parameters A and B were set to 0.99 and 0.0275 respectively (consistent with the values found in McDougle et al.[3]). The target was presented in a single location and the task comprised six 60-trial blocks, each including a 30-trial rotation phase sandwiched between a 15-trial baseline phase and 15-trial washout phase. One group experienced one of $\{\pm 15, \pm 30, \pm 45\}^\circ$in each block, and the other group experienced one of$\{\pm 45, \pm 60, \pm 90\}^\circ$, akin to experiment 1a. One participant was excluded for having a mean absolute aim of <3° (i.e., the target radius).

For Experiment 2b (n = 301) we wanted to approximate actual aiming requirements experienced by participants of a VMR study. To do this, we fit the learning rate (A) and retention rate (B) parameters of the autocorrection module to data from Bond & Taylor's eight-target reaching task.[1] Parameter fitting was carried out by finding the maximum likelihood over all participants within each rotation size separately, using the optimize.minimize function (with the 'L-BFGS-B' method) of the SciPy (scientific python) package for Python (for the 15°, 30°, 45°, 60°, and 90° rotations respectively, learning rates were 0.0385, 0.0152, 0.0152, 0.0118, 0.0044, and retention rates were 0.9985, 0.9956, 0.9951, 0.9875, 0.9918).

As in experiment 1b, the target was presented in one of eight locations, evenly spaced in 45° increments around the workspace. The order of target location was pseudorandomized, where each location occurred once in each eight-trial cycle. There was a single block of 40 cycles of rotated feedback, sandwiched between five baseline and five washout cycles. During the rotation phase, each participant experienced one rotation, sampled from $\{\pm 15, \pm 30, \pm 45, \pm 60, \pm 90\}^\circ$. Twenty-seven participants were excluded for having >20% trials with <100ms aiming time, and three participants were excluded for having a mean absolute aim of <3° (i.e., the target radius).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data Analysis

Analyses were completed with custom scripts written in Python 3.7. We used an α-value of 0.05, such that any p-value less than α indicates a statistically significant difference. Unless stated otherwise, all t-tests were two-tailed and Bonferroni corrections were applied for multiple comparisons. Our primary dependent variable was the signed difference between the aim angle and the target angle, with the resulting angle multiplied by -1 for positively signed rotations so that a positive angle was one that counteracted the applied perturbation.

Fifty-four participants were excluded (all previously listed n values were calculated before participant exclusion) because they met at least one of the following criteria: fifty participants were excluded for having >20% trials with <100ms aiming time, and four participants were excluded for having a mean absolute aim of <3° (i.e., the target radius). There was no trial-level exclusion

In experiments 1 & 2, we used t-tests to check that baseline aiming was not significantly different from zero, and to check whether aiming was significantly different from zero during mid learning and late learning. We also tested whether the washout aiming (excluding the first trial/cycle) significantly differed from zero. In experiments 1a & 1b, t-tests were used to assess whether aiming was significantly different from the ideal aim angle during mid and late learning. In experiments 2a & 2b, we instead tested if the 'final cannonball angle' (aim + offset learned by autocorrection module) was significantly different from the ideal cannonball angle during late learning and mid learning. We also applied a t-test on the implicit autocorrection vs zero during washout. In experiments 1a & 2a, mid learning was defined as the middle five rotation trials, and late learning was defined as the final five rotation trials. In experiments 1b & 2b, mid learning was defined as the middle five rotation cycles, and late learning was defined as the final five rotation cycles. In

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

experiments 1a and 2a, where participants encountered a given rotation magnitude twice with opposite signs, data was averaged over both blocks.

For the direct comparison to VMR data, we compared data from experiment 2b to data from Bond and Taylor.[1] A 2 (experiment type) x 5 (rotation size) x 3 (learning phase) ANOVA was carried out, where experiment type was denoted whether the experiment was ours or from Bond & Taylor, and learning phase was either early learning (first 10 rotation cycles), mid learning (middle 10 rotation cycles), or late learning (final 10 rotation cycles). We also applied t-tests to compare the following properties across experiments: the cycle of peak aiming, the magnitude of peak aiming, and the decrease in aiming from peak to final rotation cycle.

We then go on to investigate behavior that has been realigned to each individual's "Aha!" moment. We use paired t-tests to see if the pre-"Aha!" rotation trial has aiming that is significantly different from baseline, and we also use one-sample t-tests to test If the "Aha!" trial aiming magnitude is different from the ideal aiming magnitude. We also used t-tests to detect any significant change in trial-to-trial variability between the baseline phase and the pre-"Aha!" rotation phase, or if there was any significant error correction in the pre-"Aha!" rotation phase.

Data-driven "Aha!" moment detection

To detect the "Aha!" moment, defined as the first changepoint in aiming behavior during the rotation phase, we applied a non-parametric, model-free changepoint algorithm to each individual's aiming magnitude trial series. We used the Pruned Exact Linear Time (PELT[68]) implemented in the rupture (version 1.1.7) Python library.[69] The PELT algorithm minimizes a penalized cost function over possible segmentations. The cost function employed a radial basis function (RBF) kernel to measure dissimilarity between proposed segments, handling arbitrary shifts and the potentially multimodal nature of behavior after the "Aha!" moment. For all datasets in the main analyses, we used the same parameter settings: the penalty ('pen'), which determines the penalty of each detected changepoint (higher values lead to fewer detected changepoints), was set to 0.8; the minimum segment size ('min_size') was set to 1 (such there need not be multiple trials between each changepoint); the 'jump' parameter was set to 1, to ensure that a segmentation was considered at every trial.

The first changepoint during the rotation phase was then selected as the "Aha!" trial. In the supplementary analysis of Ding et al.[46] (Figure S1), 'pen' was set to 0.2 for greater sensitivity to changes, but all other parameters were left unchanged. Importantly, during the rotation phase, participants in Ding et al. were explicitly instructed that cursor feedback would no longer correspond to their movements and that they should discover the movement that correctly results in the cursor being placed on the target. This instruction appears to have expedited the "Aha!" moment for many participants; as such, for some analyses (e.g., Figures S1B and S1F) we exclude 145/560 participants who pre-emptively changed their aim angle on the first rotation trial (i.e., before observing any rotated feedback).

Computational models

As part of our investigation into the individual time course of aiming, we compare three models of noisy aiming: a single-process SSM to model monotonic, gradual learning; a bespoke Q-Learning model of trial-and-error learning; and the Q-Learning model embedded in a two-state HMM as the "Aha!" moment model.

State-space model

The state of the gradual-learning model, x_t, determines the aim on trial t. We assume that the process of executing an intended plan is inherently noisy, such that the aim on trial t is defined

$$a_{t,Gradual} = x_t + \epsilon \quad \text{(Equation 4)}$$

where $a_{t,Gradual}$ is the aim on trial t, x_t is the internal state of the SSM, and the execution noise ϵ follows an unbiased normal distribution

$$\epsilon = \mathcal{N}\left(\mu_c, \sigma^2_{c,Gradual}\right) \quad \text{(Equation 5)}$$

defined by $\mu_c = 0$, and execution variance $\sigma^2_{c,Gradual}$. The state is updated according to

$$x_{t+1} = Ax_t - Be_t \quad \text{(Equation 6)}$$

where A is the retention rate, B is the learning rate, and e_t is the prediction error on trial t, as defined in Equation 3. The SSM has three free parameters: A, B, and $\sigma^2_{c,Gradual}$. We also tested a dual-state SSM but found that that produced worse complexity-penalized fits than the single-state SSM defined above (Figure S2).

Trial-and-error learning model

To represent pure trial-and-error learning, we implemented a Q-learning model which learns from dense (error-based) rewards and learns sign and magnitude separately via off-policy updates. The model assumes each trial is independent, with no state transition function, and so has a one-step horizon. This statelessness is justified because feedback is immediate on each trial, and the action-reward mapping is independent of action history. The action policy is updated to minimize expected feedback errors, where the action space has been hierarchically discretized into magnitudes of 1° increments $M = \{a_0, \ldots, a_{180}\}$ and sign $SGN = \{+, -\}$. The model maintains separate value functions across trials: $Q_{mag} : M \rightarrow \mathbb{R}$ and $Q_{sgn} : SGN \rightarrow \mathbb{R}$ (both initialized to zero). The policy π is hierarchically defined via two softmax functions:

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

CellPress
OPEN ACCESS

$$\pi_{mag}(m|Q_{mag}) = \frac{\exp(\beta_{mag} Q_{mag}(m))}{\sum_{m' \in M} \exp(\beta_{mag} Q(m'))} \quad \text{(Equation 7)}$$

$$\pi_{sgn}(sgn|Q_{sgn}) = \frac{\exp(\beta_{sgn} Q_{sgn}(sgn))}{\sum_{s' \in S} \exp(\beta_{sgn} Q_{sgn}(sgn'))} \quad \text{(Equation 8)}$$

where m is the candidate magnitude, sgn is the candidate sign, β_{mag} is the inverse temperature for magnitude, and β_{sgn} is the inverse temperature for sign.

Action Selection. The model generates actions by sampling from the hierarchical mixture policy over the expanded action space:

$$p(a_{t,QLEARN}|Q) = \sum_{sgn' \in SGN} \sum_{m' \in M} \pi_{sgn}(sgn'|Q_{sgn}) \pi_{mag}(m'|Q_{mag}) \mathcal{N}\left(a_{t,QLEARN}; sgn' \cdot m', \sigma^2_{c,QLEARN}\right) \quad \text{(Equation 9)}$$

where $a_{t,QLEARN}$ is the aim on trial t and $\sigma^2_{c,QLEARN}$ is the execution variance.

Update Step. After observing feedback, the model performs off-policy Q updates using the temporal difference (TD) error. This off-policy update is analogous to counterfactual reasoning under the assumption that the action-feedback mapping is uniform across the workspace (e.g., "what would the outcome have been if I had aimed there instead?" and "where would have been the best direction to have aimed?"). Each update utilizes the maximum attainable reward for the specific sign or magnitude being updated, where reward is defined

$$r(a_{t,QLEARN}, e_t) = -|a_{t,QLEARN} + e_t| \quad \text{(Equation 10)}$$

where the prediction error e_t is counterfactually assumed to be the same for all aim angles. For sign updates, compute the maximum reward attainable $(max\ R_{sgn})$ using each sign $s \in S$:

$$max\ R_{sgn}(sgn) = \max_{m \in M} r(sgn \cdot m, e_t) \quad \text{(Equation 11)}$$

then compute the TD error δ_{sgn} and update Q_{sgn}:

$$\delta_{sgn}(sgn) = max\ R_{sgn}(sgn) - Q_{sgn}(sgn) \quad \text{(Equation 12)}$$

$$Q_{sgn}(sgn) \leftarrow Q_{sgn}(sgn) + \alpha \cdot \delta_{sgn}(sgn) \quad \text{(Equation 13)}$$

where $\alpha \in (0, 1]$ is the learning rate. A similar process is then applied to update Q_{mag}; for each magnitude $m \in M$, compute the maximum reward availalable $max\ R_{mag}$ over either sign and update with the TD error δ_{mag}:

$$max\ R_{mag}(m) = \max_{s \in S} r(sgn \cdot m, e_t) \quad \text{(Equation 14)}$$

$$\delta_{mag}(m) = max\ R_{mag} - Q_{mag}(m) \quad \text{(Equation 15)}$$

$$Q_{mag}(m) \leftarrow Q_{mag}(m) + \alpha \cdot \delta_{mag}(m) \quad \text{(Equation 16)}$$

This Q-learning model has four free parameters: $\sigma^2_{c,QLEARN}, \alpha, \beta_{mag}$, and β_{sgn}.

The "Aha!" model

We hypothesize that humans first require an "Aha!" moment to realize that a new aiming solution must be generated. To implement this, we extend the Q-learning model to be part of a two-state Hidden Markov Model (HMM), where the learner transitions between a veridical-feedback state s_v, where no learning occurs, and a perturbed-feedback state s_p, where learning is governed by the Q-learning model defined above. The HMM enables a potentially abrupt transition from s_v to s_p (and back). This transition may be arbitrarily delayed, as determined by the initial state probabilities $\mathbf{P_0} = [1, 0]$ and symmetrical prior transition matrix defined:

$$\mathrm{T} = \begin{pmatrix} 1-k & k \\ k & 1-k \end{pmatrix} \quad \text{(Equation 17)}$$

where $k \in [0, 1)$ is the prior transition probability (low values favor persistence in the current state). The prior predicted state distribution at trial t is therefore $\mathbf{P}_{t|t-1} = \mathbf{P}_{t-1}\mathrm{T}$.

Action Selection. This Q-HMM "Aha!" model first samples the hidden state $S_t \sim \mathbf{P}_{t|t-1}$, where $S_t \in \{s_{rel}, s_{irr}\}$, then an intended aim, conditionally onS_t. Specifically, for $S_t = s_{irr}$, the sensory prediction error is considered irrelevant to the agent's goals, so the planned aim is 0°. For $S_t = s_{rel}$, where the prediction error is assumed relevant to the agent's goals, the aim is sampled from the hierarchical mixture policy $a \sim \pi(\cdot|Q_{mag}, Q_{sgn})$ where $\pi(a|Q_{mag}, Q_{sgn}) = \pi_{mag}(|a||Q_{mag}) \cdot \pi_{sgn}(sgn(a)|Q_{sgn})$ using Equations 7 and 8. Thus, on trial t, the action $a_{t,\mathrm{AHA}}$ is drawn from the state-marginalized predictive distribution defined by:

Please cite this article in press as: Townsend et al., An Aha moment precedes the strategic response to a visuomotor rotation, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.04.021

$$p\left(a_{t,\mathrm{AHA}}\middle|\mathbf{P}_{\boldsymbol{t}|\boldsymbol{t}-\mathbf{1}},Q_{mag},Q_{sgn}\right)=\mathbf{P}_{\boldsymbol{t}|\boldsymbol{t}-\mathbf{1}}(s_{irr})\mathcal{N}\left(a_{t,HMM};0,\sigma^2_{c,AHA}\right)+\mathbf{P}_{\boldsymbol{t}|\boldsymbol{t}-\mathbf{1}}(s_{rel})\cdot\sum_{a'\in\mathcal{A}}\pi\left(a'\middle|Q_{mag},Q_{sgn}\right)\mathcal{N}\left(a_{t,\mathrm{AHA}};a',\sigma^2_{c,\mathrm{AHA}}\right)\quad\text{(Equation 18)}$$

where $\mathcal{A}=\{-180,\ldots,179\}$ uses the same discretized action space as the base Q-learning model, and $\sigma^2_{c,AHA}$ is execution variance.

State Inference and Update Step. The sensory prediction error (e_t) serves as the context emission that drives the posterior state distribution $\mathbf{P}_{\boldsymbol{t}}$. The state-conditioned emission probability $p(e_t|s_t)$ evaluates how well the predicted aim distribution of each state would minimize the prediction error under contextual uncertainty $\sigma^2_{context}$ (i.e., uncertainty over the prediction error):

$$p(e_t|s_t)=\begin{cases}\mathcal{N}\left(e_t;0,\sigma^2_{context}\right), & \text{if } S_t=s_v\\ \sum_{a'\in\mathcal{A}}\pi\left(a'\middle|Q_{mag},Q_{sgn}\right)\mathcal{N}\left(e_t;-a',\sigma^2_{context}\right), & \text{if } S_t=s_p\end{cases}\quad\text{(Equation 19)}$$

The posterior is then given by

$$\mathbf{P}_{\boldsymbol{t}}(S_t)=\frac{\mathbf{P}_{t|t-1}(S_t)p(e_t|S_t)}{\sum_{S'\in\{s_{rel},s_{irr}\}}\mathbf{P}_{t|t-1}(s')p(e_t|S')}\quad\text{(Equation 20)}$$

for all $S_t\in\{s_{rel},s_{irr}\}$. The Q functions ($Q_{mag},Q_{sgn}$) are updated using the off-policy Q-learning rule as in Equations 13 and 16, but conditioned on the posterior probability of being in the prediction-error-is-relevant state $\mathbf{P}_{\boldsymbol{t}}(S_t=s_{rel})$. This conditioning means that learning may stay effectively at zero until an abrupt "probability explosion" occurs (often within a single trial; Figure S2). This "probability explosion" is triggered when the posterior probability of being in state s_{rel} jumps from very small (i.e., <<1e-5) on trial *t-1* to ≈1 on trial *t* (i.e., from $\mathbf{P}_{\boldsymbol{t}}(S_t=s_{rel})\approx0$ on trial *t-1* to $\mathbf{P}_{\boldsymbol{t}}(S_t=s_{rel})\approx1$ on trial *t*). This means that the Q-functions (Q_{mag},Q_{sgn}) on trial *t* receive a full update, allowing one or both dimensions of the policy on trial *t+1* to be centered on the ideal value. This is akin to a human experiencing an "Aha!" moment, after observing feedback on trial *t*, and then implementing their new strategy on trial *t+1*. See Figure S1A for actual human trial-series aiming data overlaid on model simulations using fitted parameters.

The model is not intended to be an exhaustive mechanistic account; instead, it abstracts away some of the more ethereal components that lead up to the "Aha!" moment by overloading the state inference architecture. Put simply, the state inference architecture is perhaps a partial black-box proxy for the poorly understood processes that culminate in an "Aha!" moment.

The "Aha!" model has six free parameters: $\sigma^2_{c,\mathrm{AHA}},\sigma^2_{context},\alpha,\beta_{mag},\beta_{sgn}$, and k.

Implicit learning extension

To apply these models in tasks with implicit learning, we extend all three with the same implicit learning component defined in Equation 3. In this way, the implicitly learned bias on trial t, $\overrightarrow{x_t}(a_t)$, is added to the feedback rotation, such that the feedback on a given trial is computed:

$$EndPoint_{t,i}=a_{t,i}+Rotation_t+\overrightarrow{x_t}\left(a_{t,i}\right)+\mathcal{N}\left(0,\sigma^2_{c,i}\right)\quad\text{(Equation 21)}$$

where $i\in\{SSM,QLEARN,\mathrm{AHA}\}$ is the model being used. Each model then rolls out as previously defined with this modular implicit learning added to the feedback. This implicit module does introduce a direct dependency on past behavior, but it is negligible, so we keep the stateless nature of the Q-learning model (and its instantiation within the "Aha!" model). There are now three additional free parameters during the fitting process of each model to the Bond & Taylor dataset; the implicit learning rate A_{imp}, the implicit retention factor B_{imp}, and the standard deviation of implicit gaussian generalization σ_{imp}. When fitting to the 8-target Cannon Game-with-autocorrection dataset, these implicit parameters were fixed to the values used by the autocorrection module in the task.

All models were fit to the first block of each participant separately, for each experiment separately. Subsequent blocks were excluded as behavior there may be contaminated by experience from previous blocks. Parameters were optimized through maximum likelihood estimation using the L-BFGS-B algorithm (from the SciPy. Optimize package). Each model was fit 500 times to each participant, each time using pseudo-random initial parameter values generated through Latin hypercube sampling over the entire parameter space.