Article

Evidence accumulation from experience and observation in the cingulate cortex

https://doi.org/10.1038/s41586-025-09885-0

Received: 3 March 2025

Accepted: 7 November 2025

Published online: 07 January 2026

Check for updates

Ruidong Chen[1,2,5], Setayesh Radkani[1,2,5], Neelima Valluru[2], Seng Bum Michael Yoo[3] & Mehrdad Jazayeri[1,2,4]✉

We use our experiences to form and update beliefs about the hidden states of the world[1–3]. When possible, we also gather evidence by observing others. However, how the brain integrates experiential and observational evidence is not understood. We studied the dynamics of evidence integration in a two-player game with volatile hidden states. Both humans and monkeys successfully updated their beliefs while playing the game and observing their partner, although less effectively when observing. Electrophysiological recordings in animals revealed that the anterior cingulate cortex integrates independent sources of experiential and observational evidence into a coherent neural representation of dynamic belief about the environment's state. The geometry of population activity revealed the computational architecture of this integration and provided a neural account of the behavioural asymmetry between experiential and observational evidence accumulation. This work lays the groundwork for understanding the neural mechanisms underlying evidence accumulation in social contexts in the primate brain.

A hallmark of cognition is the ability to infer the hidden causes of our experiences. Waking up with an upset stomach, you might wonder if it is due to food poisoning or if you caught the flu. Your first clues are your symptoms—nausea versus fever would hint at different causes. But you may also rely on others' experiences. You may think it is flu if a coworker recently had the flu, or you may reason it is food poisoning if your dinner partner has similar symptoms. Although the capacity to integrate experiential and observational evidence to infer hidden causes is unequivocal, the neural mechanisms that enable such sophisticated computations are not well understood.

The anterior cingulate cortex (ACC) is thought to have a central role in evidence-based decision-making. The ACC carries signals related to outcome history, performance monitoring, action and strategy selection and beliefs about associations and contexts[1–21]. Notably, ACC representations persist over relatively long time scales[22] and integrate information across events and experiences[1–3,23]. These findings coupled with complementary causal studies[3,15,24–26] have provided strong evidence that the ACC encodes behaviourally relevant beliefs about latent causes in the environment.

We know far less about the computational and neural basis of observational inference and learning. Some studies have examined the neural signatures of observed reward and punishment[27] in the amygdala, striatum and many cortical areas[28–39]. Among these, the ACC has been a prominent region of interest that supports vicarious reinforcement and observational fear conditioning[33,35,36,39–42]. Therefore, the ACC may have a more general role in belief updating that spans both experiential and observational settings.

However, most studies on observational learning have relied on relatively simple tasks that either do not require inferring latent causes or do not involve integration of experiential and observational evidence. As a result, little is known about the parallels and distinctions between the neural representation of experiential and observational evidence and how these two sources of information are integrated to form beliefs about the latent state of the world.

Here, we tackle these questions using a combination of human behaviour, primate neurophysiology and neural modelling in a two-player belief-updating game. The behavioural results revealed a familiar asymmetry between experiential and observational evidence[43–45] in both humans and monkeys. ACC recordings revealed how the evidence derived from self and other experiences is integrated into a coherent population pattern of neural activity supporting participants' beliefs and behaviour on a trial-by-trial basis. Moreover, the organization of the population activity associated with self-experience, observation and integrated belief provides an explanation for the behavioural asymmetry.

Behavioural task and performance

We designed a two-player game for humans and monkeys to investigate the behavioural and neural signatures of updating beliefs in the presence of both experiential and observational evidence (Fig. 1a–g). Each trial consists of two phases. In the first phase, the players use their respective joysticks to independently choose between a left and a right arena (Fig. 1a). They are free to choose either arena (Fig. 1d), but this choice is consequential because only one of the arenas may eventually lead to a reward (Fig. 1c). Human participants also reported their choice confidence (Methods).

[1]Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Biomedical Engineering, Sungkyunkwan University, Suwon, Republic of Korea. [4]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]These authors contributed equally: Ruidong Chen, Setayesh Radkani. ✉e-mail: mjaz@mit.edu

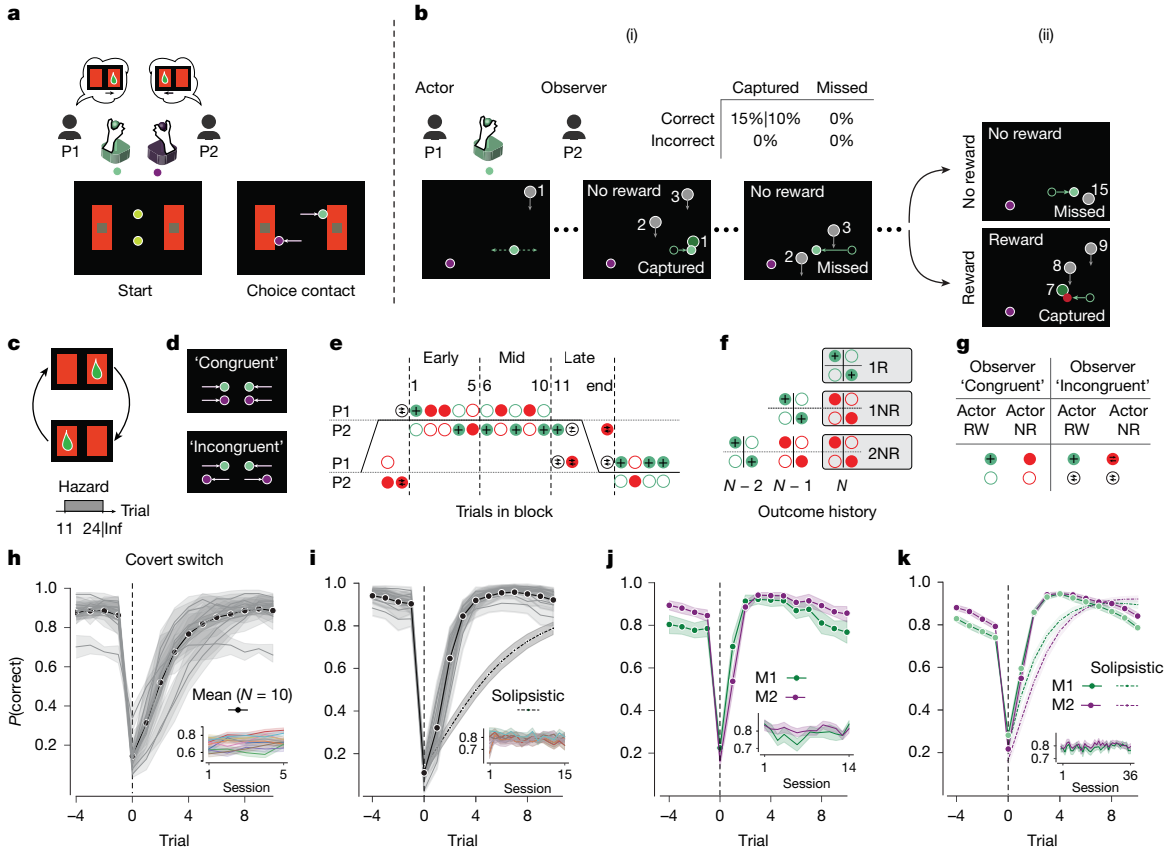

Fig. 1 | Multi-agent evidence integration task and performance. a,**b**, Trial structure. Phase 1 (**a**), players choose one of two arenas. Start: two avatars are presented in the middle of two arenas. Choice contact: when an avatar contacts an arena, the choice is registered and the avatar stops moving. Phase 2 (**b**), the actor plays a token collection game (i) for reward (ii). The observer appears below their chosen arena. The trial ends after reward or 15 unrewarded tokens. The actor receives reward probabilistically for correct choices (table). **c**, Covert switch schedule. On each trial, only one arena has a non-zero reward probability (green drop). **d**, In phase 1, players can choose the same ('Congruent') or different arenas ('Incongruent'). **e**, Example run. Grey lines denote arenas. Correct arena switches in blocks (piecewise black line). Disks show trial conditions (actor/observer, filled/open; left/right, near top/bottom horizontal line; rewarded/unrewarded, green/red; incongruent observer, opposite arrows). Numbers indicate trial in block starting at the first rewarded trial (that is, subjective block switch). Trials 1–5, 6–10 and beyond 11, demarcated by dashed lines, are Early, Mid and Late subdivisions, respectively, within a block. **f**, Outcome history nomenclature. 1R, congruent rewarded trial. 1NR, 2NR, successive congruent unrewarded trials in the same arena following 1R. **g**, Key for actor and observer and reward conditions in **e**,**f**. **h**, Single-player human performance. *P*(correct), aligned to covert block switches (trial 0). Black line, mean across participants; shade, 95% confidence interval (CI). Inset, *P*(correct) over sessions; each line corresponds to one participant. **i**, Two-player human performance. Solid, mean performance of all participants. Dashed, solipsistic model. Human performance was significantly better than the solipsistic model ($P = 2.9 \times 10^{-28}$, two-sided *t*-test). **j**,**k**, Same as **h** and **i** for one monkey (**j**) and two monkeys (**k**). Two-player performance was significantly better than the solipsistic model ($P = 4.4 \times 10^{-8}$ for M1, $P = 6.7 \times 10^{-19}$ for M2, two-sided *t*-test). Inf, infinity; P1, player 1; P2, player 2; RW, rewarded trial.

In the second phase, one player is randomly designated as the actor. The actor controls an avatar in their chosen arena with the joystick and must capture tokens falling from the top of the screen, aiming to collect as many as possible to maximize expected reward. Receiving reward is probabilistic and depends on both the arena and the number of captured tokens during the second phase (Fig. 1b). If the actor selects the correct arena, each captured token has a fixed probability of yielding a reward (0.1 for humans and 0.15 for monkeys), and the delivery of reward terminates the trial (that is, no more token capture allowed). In the other arena, capturing tokens never results in a reward. Trials without rewards end after all 15 tokens have dropped. The correct arena switches in a blocked fashion (Fig. 1c). From the moment the actor begins to collect tokens to the end of the trial, the other player, whom we refer to as the observer, watches the actor play and witnesses the outcome without receiving any reward.

We matched the actor's and observer's sensory experiences as closely as possible: both saw all events, the actor was designated only after both chose an arena, and roles were assigned randomly. Thus, both had equal information for rational inference, ensuring any asymmetries in evidence accumulation reflect internal rather than external factors.

We collected data from ten humans (five pairs) and two monkeys. All participants learned the task in a single-player version (average choice performance (mean ± s.d.) in humans, 71.85 ± 7.02% across five sessions, Fig. 1h; in monkey 1 (1) M1, 78.39 ± 3.34% and in monkey 2 (M2), 80.96 ± 1.65% across 14 sessions, Fig. 1j) before moving onto the two-player version (performance in humans 79.15 ± 3.05% over 15 sessions, Fig. 1i; in M1 78.68 ± 2.07% and M2 80.38 ± 1.73% over 36 sessions, Fig. 1k). Monkeys trained longer and faced higher reward probabilities. In both versions, performance dropped immediately after covert block switches and recovered within a few trials, indicating belief updating (Fig. 1h–k and Supplementary Fig. 2c–f). We used single-player data to simulate 'solipsistic' agents that ignored observer trials (Methods). As expected, solipsistic agents performed worse than their respective single-player source because of ignored observer trials (humans, 50.72 ± 10.90%; M1, 74.82 ± 3.27%; M2, 71.40 ± 3.09%) and critically, worse compared to the two-player sessions (Fig. 1i,k). Attention to observer trials was evident in monkeys' eye movement data, with preferential orienting of gaze towards the active arena (M1, 58.0 ± 22.5%; M2, 87.4 ± 18.7%; Supplementary Fig. 1h,k).

Humans and monkeys integrate experience rationally

To evaluate the degree to which participants played the game rationally, we compared their behaviour to that of an optimal 'oracle' model (Fig. 2). Specifically, we analysed the decision to switch on the next trial as a function of three factors: (1) history of trial outcomes (including the current trial), (2) position of the current trial in the block and (3) number of captured tokens in the current trial. To ensure a fair comparison between actor and observer, we first focus on congruent trials where both players chose the same arena.

Behavioural consequences of outcome history

Because only one arena yielded rewards, the oracle treated rewarded trials as confirmation of a correct choice and consecutive no-reward trials as accumulating evidence for a switch (star, Fig. 2a,b). The choice behaviour of both humans and monkeys showed a similar pattern: switch probability, denoted *P*(switch), was low after rewarded trials and increased monotonically with consecutive unrewarded trials (Fig. 2a,b and Supplementary Fig. 4a,b). This pattern was evident in single sessions (Fig. 2c, left); remained significant in a logistic regression analysis that accounted for outcome history, trial in block and number of captured tokens (Fig. 2c, right); and was corroborated by the pattern of human confidence reports (Supplementary Fig. 3a).

Behavioural consequences of trial position in the block

Because switches were less likely early on compared to later in the block, the oracle was more likely to switch later in the block (star, Fig. 2d,e). As neither the oracle nor the participants were aware of block switches, we inferred subjective block switches from the behaviour (Supplementary Fig. 2e,f) and registered trial position in subjective blocks (Methods). In accordance with the oracle, *P*(switch) for both humans and monkeys increased monotonically with trial position in the block (Fig. 2d,e and Supplementary Fig. 4c,d). This effect was evident in single sessions (Fig. 2f, left), remained significant in a multivariate logistic regression analysis (Fig. 2f, right) and was corroborated by the pattern of human confidence reports (Supplementary Fig. 3a).

Behavioural consequences of number of captured tokens

The probability of success increased with the number of captured tokens. Participants understood this contingency and aimed to maximize the number of captured tokens (Methods and Supplementary Fig. 2g,h). Moreover, as predicted from the oracle (star, Fig. 2g,h), participants must treat unrewarded trials with larger numbers of captured tokens as stronger evidence for a block switch. Qualitatively, this effect was evident in human participants' behaviour (Fig. 2g and Supplementary Fig. 3a) and to a lesser degree in monkeys (filled, Fig. 2h and Supplementary Fig. 4e,f). However, a more rigorous multiple regression analysis indicated that the effect was significant only in humans and in the actor condition for one of the animals (Fig. 2i and Supplementary Fig. 3b).

Behavioural consequences of the choice incongruence

We also analysed incongruent trials where participants chose opposite arenas. These trials were more frequent later in the block (Supplementary Fig. 5a,d), as expected by the higher switch probability, and were associated with lower confidence reports in humans (Supplementary Fig. 5c). Moreover, across human participants and one monkey, the actor was more likely to switch following incongruent trials (Supplementary Fig. 5b,e–g), indicating that players were influenced by each other's decisions on top of their experiences and outcomes.

Humans and monkeys discount observational evidence

By design, the oracle evaluated actor and observer trial outcomes identically (star, Fig. 2j,k). By contrast, humans and monkeys weighted observational evidence less than experienced evidence (Fig. 2j,k) even though actor and observer were randomly assigned and had identical visual experiences. This asymmetry was stronger in monkeys, possibly because the actor monkey received a juice reward. Notably, the asymmetry was evident in single sessions (Fig. 2l, left) and remained significant after accounting for other experimental factors (Fig. 2l, right). This result indicates that humans and monkeys were less responsive to unrewarded trials as an observer. Notably, this asymmetry was not due to players spending less time looking at the screen when designated as observer (Supplementary Fig. 1e–m). This asymmetry was also reflected in humans' confidence reports when they chose to stay in the same arena (Supplementary Fig. 3a).

Evidence integration in the cingulate cortex

Our behavioural results provided compelling evidence that monkeys, like humans, integrate experiential and observational evidence to infer latent state switches. To investigate the underlying neural computations, we recorded neural activity in the ACC (see Supplementary Table 1 for stereotaxic coordinates) simultaneously from the two animals.

Our recordings (M1, 31 sessions; M2, 20 sessions; simultaneous recording in 19 sessions) yielded 1,628 units (M1, 859; M2, 769). Most task-modulated neurons were sensitive to several variables, and their sensitivity could change throughout the trial (Fig. 3a–c). For example, we found mixed selectivity to trial outcome for the actor and observer (Fig. 3d,e and Supplementary Fig. 6a–d), with a large proportion sensitive to actor outcome (bootstrap test, $P < 0.05$ in 890 out of 1,628 = 54.7% of all neurons in both animals; Methods), a smaller proportion to observer outcome (bootstrap test, $P < 0.05$ in 478 out of 1,628 = 29.4% of all neurons in both animals) and a sizeable overlap between the two (299 out of 1,069 = 28.0% of outcome selective neurons). Moreover, this sensitivity changed throughout the trial, as evident from single neurons (Fig. 3f) and across the population (Fig. 3d,e, regression slopes). Notably, the alignment for outcome encoding between the actor and observer increased in the choice phase compared to the outcome phase (Fig. 3d,e; regression slope in choice, 0.46 ± 0.03; outcome, 0.13 ± 0.01; variance explained in choice, 32%; outcome, 6%). This result is consistent with a gradual integration of distinct actor- and observer-dependent responses into an identity-agnostic outcome representation.

We further analysed single neurons for evidence of outcome integration across actor and observer trials. Integration requires firing-rate modulations for the actor and observer to be in the same direction; opposite directions would counter integration. Accordingly, we restricted our analysis to 630 neurons with same-sign outcome selectivity for actor and observer (Fig. 3d,e, first and third quadrants) and quantified the difference between firing rates in the 1NR and 2NR conditions (Fig. 3g and Supplementary Fig. 6e,f). Across this population, 17.6% (111 out of 630) were more strongly modulated for 2NR compared to 1NR trials. This effect was also evident in the average firing rates of individual neurons (Fig. 3h).

Previous work has shown that ACC neurons integrate cross-trial evidence in single-player tasks that involve only experiential evidence[3,10,23]. As such, it is critical to subdivide 2NR trials and distinguish between actor–actor trials that involve only experiential evidence and the other three conditions that have at least one observer trial (actor–observer, observer–actor and observer–observer). Doing so, we found that 91.0% (101 out of 111) of neurons that featured evidence accumulation were sensitive to observer trials (Fig. 3g, inset).

A notable feature of behaviour was the asymmetry in evidence accumulation: unrewarded trials, matched in every other aspect, were weighed more strongly in actor trials compared to observer trials. We therefore asked whether firing-rate changes in neurons featuring evidence accumulation were also stronger for actor trials. Indeed,

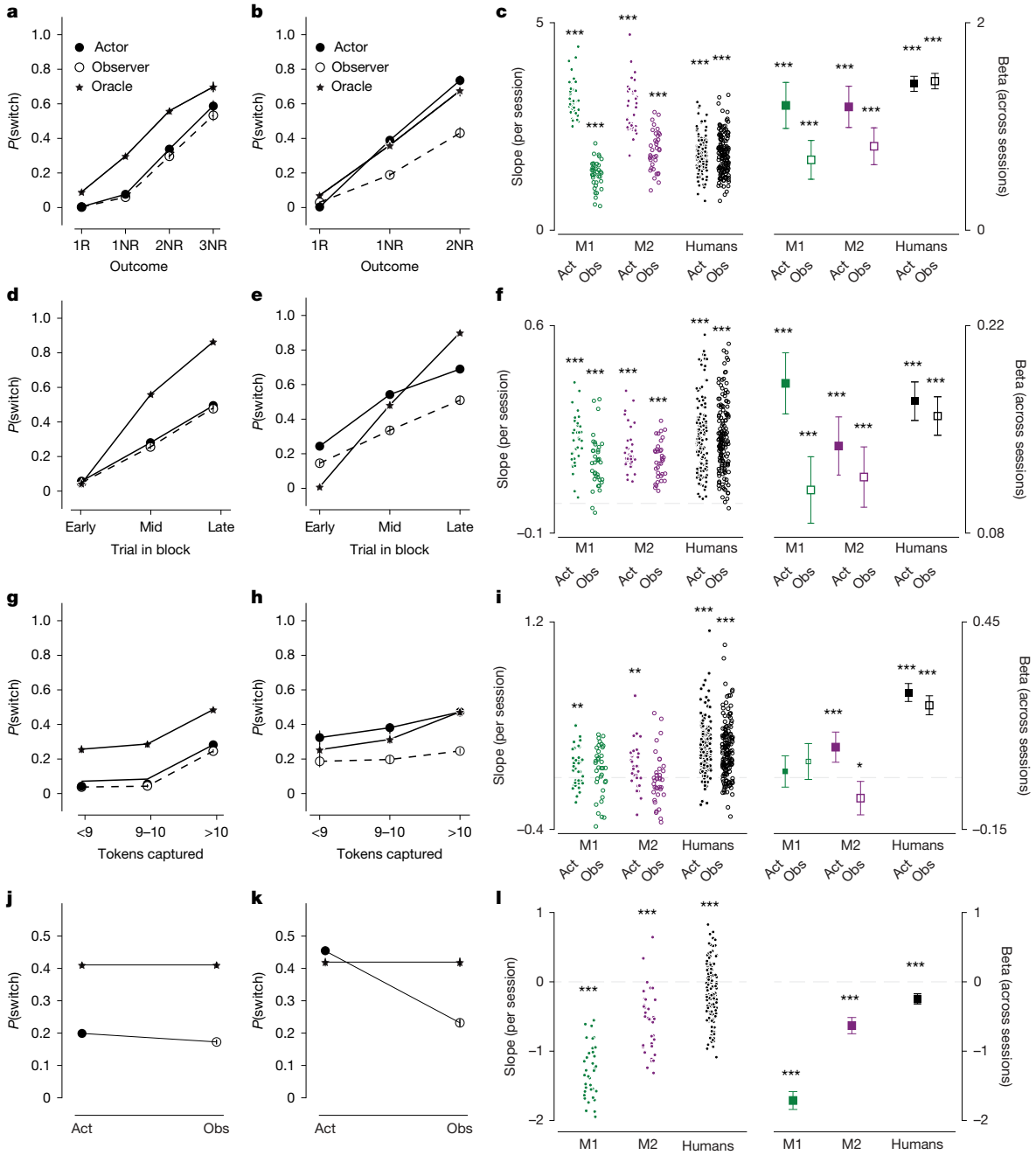

Fig. 2 | Behavioural characteristics of experiential and observational learning in monkeys and humans. a,**b**, *P*(switch) as a function of trial outcome for congruent trials across humans (**a**) and monkeys (**b**). Results are shown separately for actor (filled circle), observer (open circle) and oracle (star). *n*NR, *n*th consecutive unrewarded trial in the same arena. **c**, Left, regression slope relating *P*(switch) to outcome history, separately for actor (filled circle) and observer (open circle) for each participant for each session. $N = 36$ sessions for M1 (green) and M2 (magenta), 75 sessions for humans (5 pairs of 15 sessions each). Asterisks indicate statistical significance (two-sided *t*-test; $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$). Right, beta values of a multivariable logistic regression relating *P*(switch) to outcome history, trial in block and number of captured tokens for actor (filled square) and observer (open square) trials for each participant across sessions. Data are presented as mean values plus or minus 95% CI; asterisks indicate statistical significance (two-sided *t*-test; $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$). **d**–**f**, Same as **a**–**c** for the relationship between *P*(switch) and trial position in the block for congruent unrewarded trials: humans (**d**), monkeys (**e**) and regression slope and beta values (**f**). **g**–**i**, Same as **a**–**c** for the relationship between *P*(switch) and the number of tokens captured for congruent unrewarded trials: humans (**g**), monkeys (**h**) and regression slope and beta values (**i**). **j**–**l**, Same as **a**–**c** for the relationship between *P*(switch) and player identity (actor versus observer) for congruent unrewarded trials: humans (**j**), monkeys (**k**) and regression slope and beta values (**l**). Regressions include both actor and observer trials. Act, actor; Obs, observer.

firing-rate modulations in the 1NR and 2NR trials compared to rewarded trials were stronger in the actor condition (Fig. 3i and Supplementary Fig. 6h,i; mean *Z*-scored rate change in actor condition, 0.36 in 1NR, 0.42 in 2NR; observer condition, 0.12 in 1NR, 0.20 in 2NR; $P < 0.001$ between actor and observer conditions, paired *t*-tests).

Neural geometry of two-player evidence integration

Although single neurons encoded a wide range of task variables, this sensitivity was typically mixed, changing both during the trial (choose versus collect versus outcome phases) and as a function of trial type

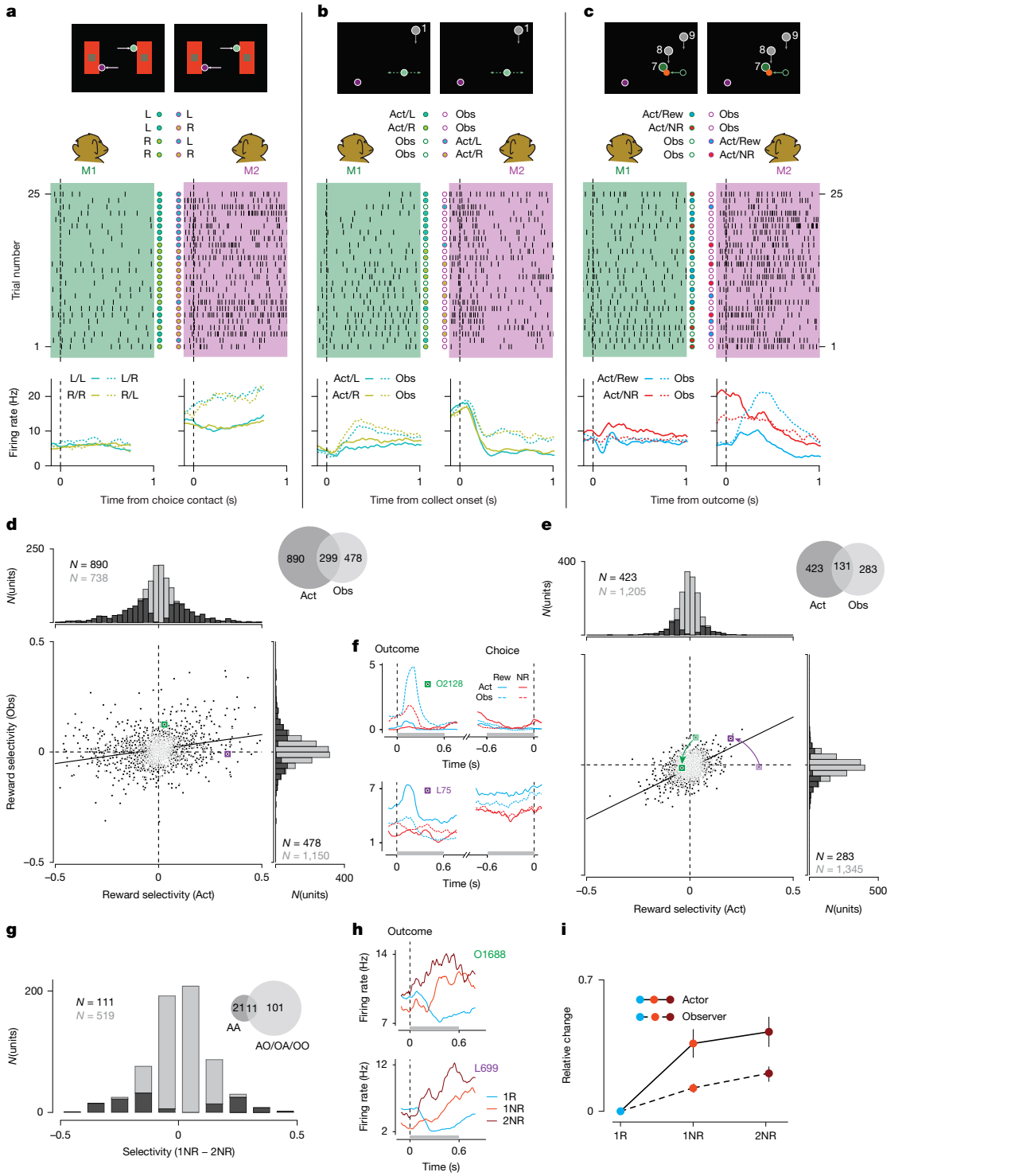

Fig. 3 | ACC neurons encode and integrate actor and observer outcome. **a**, Top, screen display at choice contact. L/R, choice of each monkey. Filled (open) circle indicates actor (observer). Middle, raster plot of action potentials recorded from a unit in M1 (left, green shading) and M2 (right, magenta shading). Dashed vertical line, choice contact. Trials are chronologically ordered from 1 to 25. Bottom, firing-rate histogram of each unit. **b**,**c**, Same as **a** with the same two units for the collect (**b**) and outcome (**c**) phases. **d**, Reward selectivity after outcome; data pooled across sessions. Scatter, actor (*x*) versus observer (*y*) selectivity (Methods). Black line, regression. Black dots indicate significant selectivity in either condition. Top left and right, histograms show significant reward selectivity (actor, 890 out of 1,628; observer, 478 out of 1,628; permutation test, 1,000 times, $P < 0.05$). Top right, number of reward selective neurons. **e**, Same as **d** before choice contact. **f**, Firing rate histograms of example neurons as shown in **d**,**e**. Top, neuron from M1. Grey bars correspond to the 600-ms windows following outcome or before choice used for analysis in **d**,**e**. Bottom, example neuron from M2. **g**, Histogram of accumulation selectivity (Methods). Black bars, neurons with significant selectivity in any of the four conditions (111 out of 630; permutation test, 1,000 times, $P < 0.05$). **h**, Firing rate histogram of example neurons with significant accumulation selectivity. Top, example neuron from M1. Dashed line aligned to outcome. Bottom, example neuron from M2. **i**, Relative change in *Z*-scored firing rate in the 600-ms window following outcome feedback between 1R and 1NR or 2NR. Solid line corresponds to actor conditions in 1NR and 2NR; dashed line corresponds to observer conditions. Data are presented as mean values plus or minus 95% CI from bootstrap; $N = 1{,}000$. L, left; R, right; Rew, reward.

(for example, actor versus observer). This property, which is common across frontal cortical neurons[46], motivated further neural analysis at the population level, which can offer complementary computational insights[47–49].

A stable dimension encoding switch belief

First, we identified a dimension along which activity increased monotonically across trials leading to a behavioural switch (Fig. 4a; Methods). Using cross-validation, we verified that this dimension predicted switch behaviour (Fig. 4b; Supplementary Fig. 7a,b for individual animals). Importantly, this effect was specific to trials preceding switches and was not due to a trial-order effect (Fig. 4b, dashed). We also confirmed the link between the encoding dimension and switch behaviour by verifying that large and small projections of neural activity on the encoding dimension corresponded to high and low values of P(switch), respectively (Fig. 4c).

Because behaviour was influenced by both actor and observer trials, we examined projections onto the encoding dimension for each trial type separately. This dimension carried outcome information for both, with larger projections in 2NR than 1NR trials ($P < 0.001$ in both actor and observer conditions, paired t-tests; Fig. 4d; Supplementary Fig. 7e,f for individual animals). This observation was evident within sessions as quantified by the regression slopes relating projections to the number of unrewarded trials (Fig. 4d, inset). Consistent with behavioural asymmetry, regression slopes were steeper for actor than observer trials ($P < 0.001$, paired t-test, Fig. 4d, inset). Control analyses confirmed that this difference was not explained by unequal trial counts or neuron numbers (Supplementary Fig. 11a–c).

Identity and outcome inputs to the ACC

We examined the neural state space to characterize how inputs to the ACC convey information about identity and outcome. We considered two hypotheses. H1 proposes a shared, identity-agnostic input for outcomes and a separate input for identity. This arrangement yields parallel encoding dimensions for actor and observer outcomes (Fig. 4e). H2, in contrast, posits independent inputs for experienced and observed outcomes, producing orthogonal encoding dimensions (Fig. 4i).

To test these hypotheses, we compared the geometry of ACC population activity with that of recurrent neural network (RNN) models implementing each hypothesis. In H1, the RNN received one input conveying identity-agnostic outcome information and another signalling identity (Fig. 4f). In H2, it received separate inputs for experienced and observed outcomes (Fig. 4j). We trained 100 randomly initialized RNNs per hypothesis (Methods).

All RNNs successfully integrated evidence and reproduced the behavioural asymmetry (Fig. 4g,k and Supplementary Fig. 8). We then compared their outcome representational geometry to that of the ACC, identifying actor and observer encoding dimensions in each dataset and evaluating H1 and H2 on the basis of the angle between these dimensions in the models and the ACC.

We compared RNNs instantiating H1 and H2. The angle between actor and observer dimensions was smaller and closer to parallel in H1 (30.77 ± 6.88°, mean ± s.d., $N = 100$; Fig. 4h) and larger, approaching orthogonality, in H2 (78.89 ± 7.56°, $N = 100$; Fig. 4l). ACC data showed similarly large angles (M1, 90.83 ± 1.35°; M2, 75.97 ± 1.28°; $N = 100$ splits per animal), consistent with H2 predictions (Fig. 4h,l). This orthogonality was not explained by firing-rate differences in rewarded trials; differences during unrewarded trials were equally strong and frequent (rewarded, $n = 375$ out of 1,628, $P < 0.05$, average Z-scored rate difference = 0.31; unrewarded, $n = 358$ out of 1,628, average Z-scored rate difference = 0.34; Supplementary Fig. 12c). Moreover, orthogonality persisted when considering only unrewarded trials (Supplementary Fig. 12f).

To test whether the independent inputs are necessary for orthogonality, we trained further RNNs with the H1 input architecture but imposed a constraint enforcing a large angle between the actor and observer outcome dimensions. These models failed to perform the task (Supplementary Fig. 13). These findings strengthen the hypothesis that the ACC receives actor and observer outcome information by means of independent input pathways.

Input projection patterns onto the ACC

We analysed population responses in the ACC to dissect the organization of input projections onto the ACC. Our previous analysis indicated that the ACC receives independent inputs associated with actor and observer outcomes. This orthogonality is consistent with two hypotheses, denoted H2a and H2b. H2a posits that the actor and observer inputs project to disjoint ACC subpopulations (Fig. 4m and Supplementary Fig. 9a). This organization is consistent with the input orthogonality because disjoint populations are inherently orthogonal. H2b, in contrast, posits mixed projections to the same population in the ACC (Fig. 4m and Supplementary Fig. 9a), which could also result in orthogonality.

Many neurons were sensitive to both actor and observer outcomes (Fig. 3d and Supplementary Fig. 6a,b). This result provides evidence for some level of mixed projection. To distinguish between H2a and H2b more definitively, we divided the neurons into two groups. The first group, which we refer to as aligned, include neurons whose firing rates move in the same direction for actor and observer conditions, either positively or negatively (Fig. 4m and Supplementary Fig. 9a, green). The second group, which we refer to as anti-aligned, are the neurons that encode outcome with opposite signs (Fig. 4m and Supplementary Fig. 9a, magenta).

Critically, H2a and H2b make distinct predictions about the geometry of activity within the subspaces formed by these groups. In H2a, where projections are disjoint, the angles between actor and observer outcome representations remain orthogonal for both aligned and anti-aligned neurons (Fig. 4n and Supplementary Fig. 9b,c). By contrast, H2b predicts a divergence in the angle relationships: for aligned neurons, the encoding dimensions are more aligned, resulting in acute angles, whereas for anti-aligned neurons, the encoding dimensions are oppositely oriented, leading to obtuse angles (Fig. 4n and Supplementary Fig. 9b,c).

In the ACC, at the time of outcome, the decomposed angle between actor and observer dimensions had large positive and negative components, consistent with H2b—not H2a (Fig. 4o; Supplementary Fig. 7g,h for individual animals). As the task proceeds to time of choice for the next trial, the two vectors become more aligned (angle at outcome, 83.46° ± 0.92°; choice, 69.94° ± 1.98°; Fig. 4o), consistent with the increased correlation of single-neuron selectivities (Fig. 3e). Together these results indicate that the actor and observer outcome inputs drive the ACC through mixed projection patterns.

Geometry of evidence integration

For effective integration, the switch evidence (SE) dimension should be outside the null space of both actor and observer outcome encoding dimensions. This predicts that the angle between the actor outcome and SE, $\theta(O_{Act},SE)$, as well as the angle between the observer outcome and SE, $\theta(O_{Obs},SE)$ must be less than 90°. However, the relative geometry of these dimensions differ under the two hypotheses (Fig. 4p). Under H1, because the outcome encoding dimensions are parallel, $\theta(O_{Act},SE)$ and $\theta(O_{Obs},SE)$ must be the same (Fig. 4p). By contrast, under H2, because the outcome encoding dimensions are orthogonal, $\theta(O_{Act},SE)$ and $\theta(O_{Obs},SE)$ can assume different values (Fig. 4q).

We measured $\theta(O_{Act},SE)$ and $\theta(O_{Obs},SE)$ in both the models and the ACC (Methods). The representational geometry in the ACC was better captured by H2 (Fig. 4r,s). In both the ACC and H2-instantiating RNNs, $\theta(O_{Act},SE)$ was smaller than $\theta(O_{Obs},SE)$ (RNN_{H2}, 43.15° ± 10.22° versus 59.11° ± 10.89°; M1, 26.25° ± 0.88° versus 79.05° ± 1.42°; M2, 31.96° ± 0.94° versus 54.40° ± 1.35°). This difference was not explained

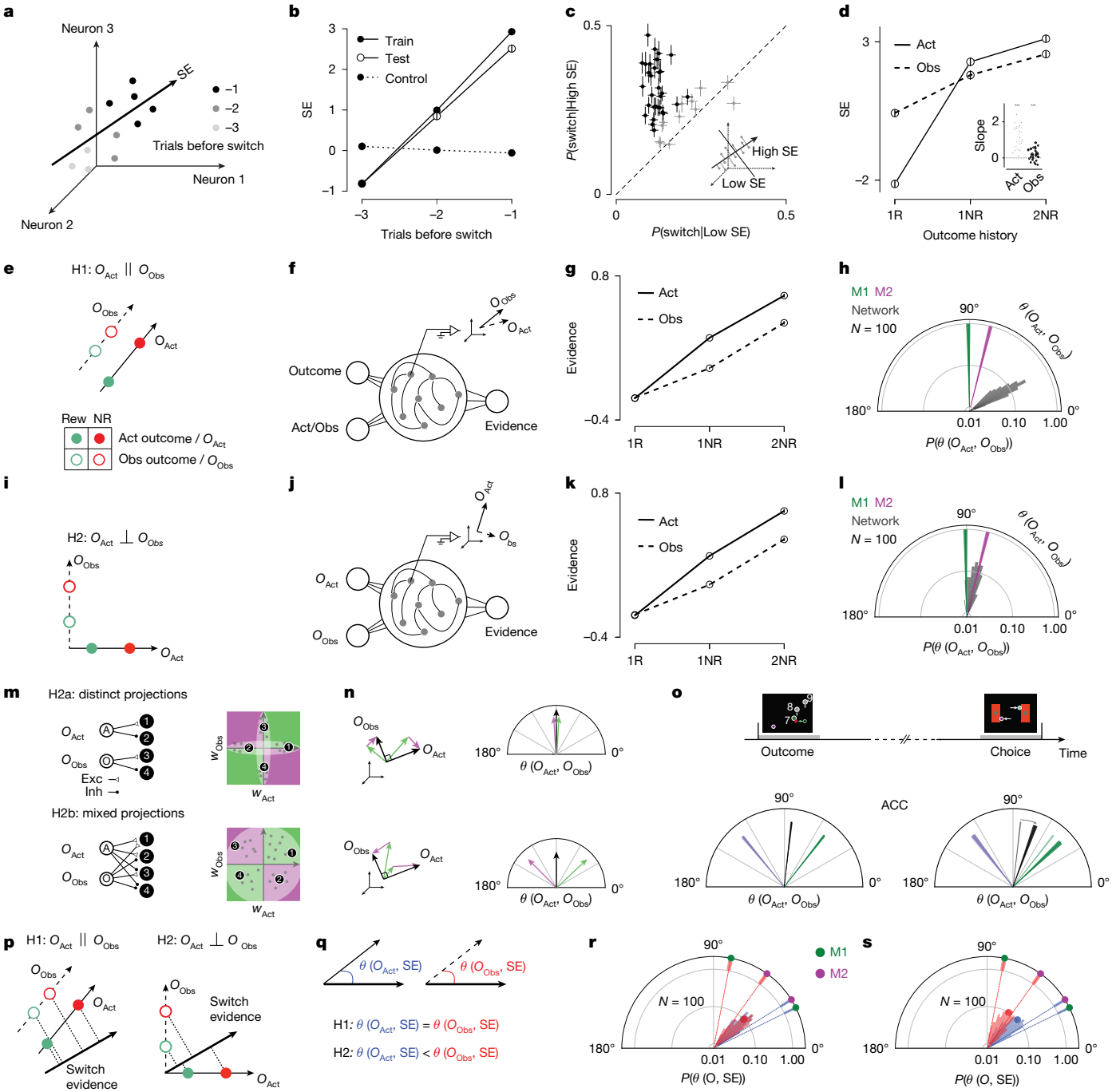

Fig. 4 | Population geometry of multi-agent evidence integration. a, SE from regression on trials before switch. **b**, Projection on SE relative to trials before switch for training (solid circle), test (open circle) and null control (dashed line). **c**, Proportion of switch trials conditioned on a low (*x* axis) or high (*y* axis) projection on SE for each session. Error bars, s.e.m. Black dots, significant differences ($N = 31/43$ sessions, rank-sum test, $P < 0.05$). Inset, trials are sorted by SE projection relative to the session median. **d**, Projection on SE by behavioural condition. Solid (dashed) line, actor (observer) conditions. Inset, slopes of regression for projection on SE over outcome history. **e**, Parallel hypothesis on the organization of neural activity. **f**, RNN instantiating the parallel hypothesis. **g**, Performance of trained RNNs in **f**. Solid (dashed) line represents output in actor (observer) condition. Error bar, 95% CI. **h**, Angle between actor and observer outcome dimensions ($\theta(O_{Act}, O_{Obs})$). Grey, from network activation over 100 model instantiations. Green (magenta), from ACC neural activity in M1 (M2) over 100 randomly selected halves of trials; width indicates s.d. **i**–**l**, same as **e**–**h** for the orthogonal hypothesis (**i**). Model instantiation (**j**); model performance (**k**); model comparison (**l**). **m**, Network architecture hypotheses. Left, schematic of weights between input populations representing actor and observer outcome and ACC neural population. Right, input weights for one neuron. **n**, Outcome dimensions (black) decomposed into aligned (green) and anti-aligned (magenta) subspaces. **o**, Decomposition of $\theta(O_{Act}, O_{Obs})$ in ACC. **p**, Geometry of outcome and SE integration axes under H1 and H2. **q**, Angular relationship between outcome and SE integration axes under H1 and H2. **r**, Centre blue/red angular distribution shows values of $\theta(O_{Act}, SE)/\theta(O_{Obs}, SE)$ for H1-instantiating RNNs. Blue/red circles, mean values. Dashed lines, ACC data; separately for the two animals (M1, green; M2, magenta). Shaded area indicates 1 s.d. computed from randomly selected halves of trials, repeated 100 times. **s**, Same as **r** for H2, with network data from **j**.

by unequal trial counts or numbers of outcome-responsive neurons and persisted after controlling for both factors (M1, 34.07° ± 1.82° versus 74.05° ± 2.44°; M2, 33.20° ± 1.37° versus 48.88° ± 1.75°; Supplementary Fig. 11d). By contrast, the two angles had similar magnitudes in the H1-instantiating RNNs (44.03° ± 7.35° versus 43.48° ± 7.80°).

For integration to occur, actor and observer outcomes must drive a common SE. In this respect, the smaller $\theta(O_{Act}, SE)$ relative to $\theta(O_{Obs}, SE)$ may provide a neural explanation of the behavioural asymmetry in experiential versus observational evidence integration. Specifically, if integration occurs by linear projection of activity along the outcome dimension onto the SE dimension, then a smaller angle between the actor outcome and SE dimensions would enable the same strength of evidence collected on actor trials to produce a higher increment in cumulative evidence (Fig. 4p).

To further substantiate the relationship between neural geometry and behavioural sensitivity, we computed the outcome-switch angle separately for each session, for both actor and observer conditions. Pooling the data across the two animals, we found strong support for our hypothesis: there was a significant negative correlation between behavioural sensitivity and neural angle across sessions (Supplementary Fig. 14a–c).

Further analyses of these variables for the two animals and the two conditions separately (Supplementary Fig. 14d–i) revealed that the effects were significant for the actor condition in M1 and the observer condition in M2 only. On the basis of this finding, we hypothesized that the two animals were engaged in a leader–follower dynamic, with M2 being more strongly influenced by M1 than the other way around, which was borne out of further behavioural analyses (Supplementary Fig. 14j; also see Fig. 2l).

Together, these results indicate that the angles between SE and outcome dimensions in the ACC account for the asymmetry in behavioural sensitivity between actor and observer conditions.

So far, we configured all RNNs such that their output weights were fixed. To test the effect of this assumption on our findings, we performed control analyses on RNNs built with learnable output weights (Supplementary Fig. 10b,d). The geometry of evidence integration in these readout-learnable networks was different from both the readout-fixed networks and the ACC. Specifically, they exhibited relatively higher alignment between actor and observer outcome dimensions and outcome and evidence dimensions (Supplementary Fig. 10b,d). These results indicate that ACC internal dynamics and not its downstream projections are responsible for evidence integration.

Discussion

Our work brings together two important yet traditionally distinct areas of research concerning the role of the ACC in cognition. One important function of the ACC is to monitor and integrate one's experience over time to inform strategic decision-making[6,11,17]. This function supports a wide range of mental computations including explore–exploit trade-offs, cost–benefit analysis, conflict monitoring and causal inference[1–4,19,23,25]. Another function ascribed to the ACC is sensitivity to observed reward and punishment, enabling vicarious learning[35,36,39–42]. Our work offers a unifying perspective wherein the ACC has a general role in integrating experiential and observational outcomes over flexible timescales to update belief about environmental states. The confluence of these two research directions brings to focus several important questions.

First, what anatomical substrates and circuit motifs enable the integration of information about self and other? We found that the ACC encodes all three key computational variables needed for integration: actor outcome, observer outcome and integrated belief. A comparison of population activity between models and the ACC provided evidence that actor and observer outcomes were associated with activity patterns in orthogonal subspaces. This finding indicates that the ACC computes beliefs about the state of the environment by integrating outcome information from distinct identity-dependent input streams.

Analysis of the geometry of neural representation is often used to infer computational algorithms. For example, subspace orthogonality is thought to prevent interference and maximize robustness[47,50–52], and factorized representations are thought to facilitate structural generalization[49,53–57]. In our work, we augmented this analysis with single-neuron tuning properties to dissect the organization of input projections onto the ACC. Results indicated that the ACC did not rely on disjoint subpopulations for actor and observer information. Instead, actor and observer information was supplied by means of overlapping projections. We do not know the constraints that determine the organization of these projections. However, in our experiment, this mixing may facilitate the integration process. With disjoint subpopulations, the integration would have to be augmented by a gating mechanism to select the subpopulation that has to be integrated on each trial. By contrast, the mixed representation provides a single subpopulation of outcome-aligned neurons that can be used for integration in all trials.

Our analyses indicated that the ACC coding properties changed throughout the inter-trial interval[58] (Fig. 4o and Supplementary Fig. 15). One notable feature was the reduction of the angle between the population vectors encoding the actor and observer outcomes, from the outcome phase to the choice phase in the next trial. This finding is reminiscent of previous work showing that high-dimensional firing-rate vectors rapidly decay to a single dimension during the process of decision-making[59]. However, in our work, this process unfolded in the presence of multiple inputs (two agents) and long timescales (across trials), which pose important constraints on the circuits responsible for evidence accumulation in ACC[60,61].

Second, does the brain process experiential and observational information similarly? In our two-player game, although humans and monkeys integrated experiential and observational outcomes, they learned less from observations. Discounting observational evidence has been reported previously[43–45]. However, several aspects of our study reinforce the view that there is a fundamental asymmetry between learning from experience and observation. By interleaving actor and observer trials while collecting each player's choice on every trial, we could track both players' evolving beliefs with precision. This design choice as well as our analysis of congruent and unrewarded trials enabled us to rule out various confounds that could lead to this asymmetry. Finally, we found this asymmetry to be stronger in monkeys, possibly because monkeys received a juice reward, which could accentuate the difference between experience and observation. The difference between humans and monkeys may also stem from superior social cognition in humans enabling more effective evaluation and integration of observations.

We identified a neural correlate of this asymmetry in the ACC, where signals encoding actor outcomes were more closely aligned with cumulative SE than those encoding observer outcomes. Validating the functional relevance of this finding will require precise patterned activations of subpopulations of neurons in the ACC[62–64]. Additionally, characterizing the behavioural contingencies and neural constraints that give rise to this asymmetry remains an important direction for future research.

In sum, our work establishes the basic mechanisms of multi-agent evidence integration and offers a starting point for addressing exciting and unresolved questions about social learning. Extensions of our work can be used to study the mechanisms through which cognitive factors such as belief about the partner's skill level, their previous knowledge about task contingencies and their social rank influence observational learning.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09885-0.

1. Hayden, B. Y., Pearson, J. M. & Platt, M. L. Neuronal basis of sequential foraging decisions in a patchy environment. *Nat. Neurosci.* **14**, 933–939 (2011).
2. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013).
3. Sarafyazd, M. & Jazayeri, M. Hierarchical reasoning by neural circuits in the frontal cortex. *Science* **364**, eaav8911 (2019).
4. Carter, C. et al. Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* **280**, 747–749 (1998).
5. Ito, S., Stuphorn, V., Brown, J. W. & Schall, J. D. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* **302**, 120–122 (2003).
6. Seo, H. & Lee, D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* **27**, 8366–8377 (2007).
7. Monosov, I. E. Anterior cingulate is a source of valence-specific information about value and uncertainty. *Nat. Commun.* **8**, 1–12 (2017).
8. Kennerley, S. W., Behrens, T. E. J. & Wallis, J. D. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat. Neurosci.* **14**, 1581–1589 (2011).
9. Kennerley, S. W. & Wallis, J. D. Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. *Eur. J. Neurosci.* **29**, 2061–2073 (2009).
10. Shima, K. & Tanji, J. Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* **282**, 1335–1338 (1998).
11. Amiez, C., Joseph, J.-P. & Procyk, E. Anterior cingulate error-related activity is modulated by predicted reward. *Eur. J. Neurosci.* **21**, 3447–3452 (2005).
12. Hadland, K. A., Rushworth, M. F. S., Gaffan, D. & Passingham, R. E. The anterior cingulate and reward-guided selection of actions. *J. Neurophysiol.* **89**, 1161–1164 (2003).
13. Williams, Z. M., Bush, G., Rauch, S. L., Cosgrove, G. R. & Eskandar, E. N. Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat. Neurosci.* **7**, 1370–1375 (2004).
14. Rushworth, M. F. S. & Behrens, T. E. J. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* **11**, 389–397 (2008).
15. Akam, T. et al. The anterior cingulate cortex predicts future states to mediate model-based action selection. *Neuron* **109**, 149–163 (2021).
16. Vertechi, P. et al. Inference-based decisions in a hidden state foraging task: differential contributions of prefrontal cortical areas. *Neuron* **106**, 166–176 (2020).
17. Shidara, M. & Richmond, B. J. Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science* **296**, 1709–1711 (2002).
18. Narayanan, N. S., Cavanagh, J. F., Frank, M. J. & Laubach, M. Common medial frontal mechanisms of adaptive control in humans and rodents. *Nat. Neurosci.* **16**, 1888–1895 (2013).
19. Quilodran, R., Rothé, M. & Procyk, E. Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* **57**, 314–325 (2008).
20. Tervo, D. G. R. et al. The anterior cingulate cortex directs exploration of alternative strategies. *Neuron* **109**, 1876–1887 (2021).
21. Tervo, D. G. R. et al. Behavioral variability through stochastic choice and its gating by anterior cingulate cortex. *Cell* **159**, 21–32 (2014).
22. Bernacchia, A., Seo, H., Lee, D. & Wang, X.-J. A reservoir of time constants for memory traces in cortical neurons. *Nat. Neurosci.* **14**, 366–372 (2011).
23. Kerns, J. G. et al. Anterior cingulate conflict monitoring and adjustments in control. *Science* **303**, 1023–1026 (2004).
24. Gu, B.-M. et al. Neural correlates of cognitive inflexibility during task-switching in obsessive-compulsive disorder. *Brain* **131**, 155–164 (2008).
25. Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J. & Rushworth, M. F. S. Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* **9**, 940–947 (2006).
26. Chen, W., Liang, J., Wu, Q. & Han, Y. Anterior cingulate cortex provides the neural substrates for feedback-driven iteration of decision and value representation. *Nat. Commun.* **15**, 6020 (2024).
27. Joiner, J., Piva, M., Turrin, C. & Chang, S. W. C. Social learning through prediction error in the brain. *NPJ Sci. Learn.* **2**, 8 (2017).
28. Grabenhorst, F., Báez-Mendoza, R., Genest, W., Deco, G. & Schultz, W. Primate amygdala neurons simulate decision processes of social partners. *Cell* **177**, 986–998 (2019).
29. Báez-Mendoza, R., Harris, C. J. & Schultz, W. Activity of striatal neurons reflects social action and own reward. *Proc. Natl Acad. Sci. USA* **110**, 16634–16639 (2013).
30. Burke, C. J., Tobler, P. N., Baddeley, M. & Schultz, W. Neural mechanisms of observational learning. *Proc. Natl Acad. Sci. USA* **107**, 14431–14436 (2010).
31. Cooper, J. C., Dunne, S., Furey, T. & O'Doherty, J. P. Human dorsal striatum encodes prediction errors during observational learning of instrumental actions. *J. Cogn. Neurosci.* **24**, 106–118 (2012).
32. Azzi, J. C. B., Sirigu, A. & Duhamel, J.-R. Modulation of value representation by social context in the primate orbitofrontal cortex. *Proc. Natl Acad. Sci. USA* **109**, 2126–2131 (2012).
33. Jeon, D. et al. Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. *Nat. Neurosci.* **13**, 482–488 (2010).
34. Gariépy, J.-F. et al. Social learning in humans and other animals. *Front. Neurosci.* **8**, 58 (2014).
35. de Araujo, M. F. P. et al. Neuronal activity of the anterior cingulate cortex during an observation-based decision making task in monkeys. *Behav. Brain Res.* **230**, 48–61 (2012).
36. Hill, M. R., Boorman, E. D. & Fried, I. Observational learning computations in neurons of the human anterior cingulate cortex. *Nat. Commun.* **7**, 12722 (2016).
37. Allsop, S. A. et al. Corticoamygdala transfer of socially derived information gates observational learning. *Cell* **173**, 1329–1342 (2018).
38. Huang, Z. et al. Ventromedial prefrontal neurons represent self-states shaped by vicarious fear in male mice. *Nat. Commun.* **14**, 3458 (2023).
39. Yoshida, K., Saito, N., Iriki, A. & Isoda, M. Social error monitoring in macaque frontal cortex. *Nat. Neurosci.* **15**, 1307–1312 (2012).
40. Chang, S. W. C., Gariépy, J.-F. & Platt, M. L. Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* **16**, 243–250 (2013).
41. Basile, B. M., Schafroth, J. L., Karaskiewicz, C. L., Chang, S. W. C. & Murray, E. A. The anterior cingulate cortex is necessary for forming prosocial preferences from vicarious reinforcement in monkeys. *PLoS Biol.* **18**, e3000677 (2020).
42. Hayden, B. Y., Pearson, J. M. & Platt, M. L. Fictive reward signals in the anterior cingulate cortex. *Science* **324**, 948–950 (2009).
43. Bellebaum, C., Jokisch, D., Gizewski, E. R., Forsting, M. & Daum, I. The neural coding of expected and unexpected monetary performance outcomes: dissociations between active and observational learning. *Behav. Brain Res* **227**, 241–251 (2012).
44. Morin, O., Jacquet, P. O., Vaesen, K. & Acerbi, A. Social information use and social information waste. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200052 (2021).
45. Nicolle, A., Symmonds, M. & Dolan, R. J. Optimistic biases in observational learning of value. *Cognition* **119**, 394–402 (2011).
46. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
47. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
48. Ebitz, R. B. & Hayden, B. Y. The population doctrine in cognitive neuroscience. *Neuron* **109**, 3055–3068 (2021).
49. Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M. & Cunningham, J. P. Reorganization between preparatory and movement population responses in motor cortex. *Nat. Commun.* **7**, 13239 (2016).
50. Johnston, W. J., Fine, J. M., Yoo, S. B. M., Ebitz, R. B. & Hayden, B. Y. Semi-orthogonal subspaces for value mediate a binding and generalization trade-off. *Nat. Neurosci.* **27**, 2218–2230 (2024).
51. Flesch, T., Juechems, K., Dumbalska, T., Saxe, A. & Summerfield, C. Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron* **110**, 4212–4219 (2022).
52. Tang, C., Herikstad, R., Parthasarathy, A., Libedinsky, C. & Yen, S.-C. Minimally dependent activity subspaces for working memory and motor preparation in the lateral prefrontal cortex. *eLife* **9**, e58154 (2020).
53. Bernardi, S. et al. The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* **183**, 954–967 (2020).
54. Remington, E. D., Narain, D., Hosseini, E. A. & Jazayeri, M. Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron* **98**, 1005–1019 (2018).
55. Lindsey, J. W. & Issa, E. B. Factorized visual representations in the primate visual system and deep neural networks. *eLife* **13**, RP91685 (2024).
56. Ito, T. et al. Compositional generalization through abstract representations in human and artificial neural networks. In *Proc. 36th International Conference on Neural Information Processing Systems* (eds Koyejo, S. et al.) 32225–32239 (ACM, 2022).
57. Johnston, W. J. & Fusi, S. Abstract representations emerge naturally in neural networks trained to perform multiple tasks. *Nat. Commun.* **14**, 1040 (2023).
58. Daie, K., Fontolan, L., Druckmann, S. & Svoboda, K. Feedforward amplification in recurrent networks underlies paradoxical neural coding. Preprint at *bioRxiv* https://doi.org/10.1101/2023.08.04.552026 (2023).
59. Ganguli, S. et al. One-dimensional dynamics of attention and decision making in LIP. *Neuron* **58**, 15–25 (2008).
60. Brown, J. W. & Alexander, W. H. Foraging value, risk avoidance, and multiple control signals: how the anterior cingulate cortex controls value-based decision-making. *J. Cogn. Neurosci.* **29**, 1656–1673 (2017).
61. Vassena, E., Holroyd, C. B. & Alexander, W. H. Computational models of anterior cingulate cortex: at the crossroads between prediction and effort. *Front. Neurosci.* **11**, 316 (2017).
62. Mardinly, A. R. et al. Precise multimodal optical control of neural ensemble activity. *Nat. Neurosci.* **21**, 881–893 (2018).
63. Clark, A. M. et al. An optrode array for spatiotemporally-precise large-scale optogenetic stimulation of deep cortical layers in non-human primates. *Commun. Biol.* **7**, 329 (2024).
64. Russell, L. E. et al. All-optical interrogation of neural circuits in behaving mice. *Nat. Protoc.* **17**, 1579–1620 (2022).

Methods

We collected behavioural data from humans and behavioural and neurophysiological data from rhesus macaque monkeys (*Macaca mulatta*). Experimental procedures for humans were approved by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology. Experimental procedures for animals conformed to the National Institutes of Health guidelines and were approved by the Committee of Animal Care at the Massachusetts Institute of Technology.

Experimental procedures for non-human primates

Two monkeys (M1, female, 6 kg, aged 6; M2, male, 11 kg, aged 11) were seated comfortably in two adjacent primate chairs in a dark, quiet enclosure at a distance of 40 inches. Animals were head-restrained, facing forwards and unable to see one another. Stimuli were presented on two side-by-side display monitors (Acer R240HY) 40 inches apart (centre to centre), at a normal distance of 19 inches from the animals' eyes. The monitors displayed identical stimuli throughout experiments. Each animal could manipulate a joystick (Logitech Extreme 3D Pro) placed at a distance adjusted for each animal's reach (about 6 inches) in front of their chair. Joysticks were physically constrained to left and right movements only. The joystick digital output (1–1,024) was thresholded to three states: left movement (1–463), no movement (464–560) and right movement (561–1,024). Eye movements were sampled at 1 kHz using infrared cameras (Eyelink 1,000, SR Research). The MWorks software package (http://mworks-project.org) and MOOG library[65] (https://jazlab.github.io/moog.github.io/) were used to generate visual stimuli and to enforce behavioural contingencies. A photodiode was used to sync electronic events with stimulus presentations.

Neural recordings were made from the ACC with 64-channel linear probes with 50-μm interelectrode spacing (V-probe, Plexon Inc.) inserted through a rectangular recording chamber. Extracellular signals were bandpass filtered (300 Hz to 6 kHz) and digitized (sampling rate, 30 kHz) using two 32-channel headstages (Intan Technologies) and collected using OpenEphys software (http://www.open-ephys.org). Spike sorting and curation were carried out using Kilosort 3 (https://github.com/MouseLand/Kilosort) and phy (https://github.com/cortex-lab/phy). Recording sites and number of sessions/trials are reported in Supplementary Table 1. Data analysis was performed using custom Python code.

Experimental procedures for humans

We recruited a total of 14 participants. All participants gave informed consent, were naive to the purpose of the study and had normal or corrected-to-normal vision. Participants were asked to play the single-player version of the task first and after gaining familiarity were invited to play the two-player version. One participant did not learn the task after three single-player sessions (set to be an exclusion criterion before data collection began). Three participants withdrew voluntarily after finishing the single-player sessions. The remaining ten participants (six men and four women, aged 18–65) completed the two-player sessions. These participants were divided into five fixed pairs (one female–female, two male–male and two male–female). Each session lasted roughly 60 min. Participants were paid a fixed amount at the end of each session. Those who finished all single- and two-player experiments were paid a further 50% of all their earnings as a bonus.

In each session, participants sat in adjacent dark enclosures, separated by an opaque curtain to prevent visual contact. Each enclosure had identical equipment (monitor, keyboard, joystick connected to a Mac mini). Participants knew their monitors displayed the same visuals, although they could not see the other screen. In the single-player experiment, only one setup was used. As in the monkey experiments, tasks, stimuli and behavioural contingencies were controlled by MWorks and MOOG software packages.

Behavioural task

We devised a two-player trial-based game. Each trial consists of two hierarchically organized phases. Phase 1 begins with the presentation of the two arenas on the two sides of the monitor and two disks stacked on top of one another midway between the two arenas representing the two players ('avatars'). Each arena is a red rectangle (8.2 cm × 16.4 cm) with a central grey square (width, 4.1 cm) placed 3 cm to the left or right of the vertical midline. Each avatar appears as a yellow disk (diameter, 2.4 cm) placed 3.85 cm above or below the horizontal midline. After 750 ms, the two yellow avatars are randomly assigned to the two players by a change of colour, purple for player 1 and green for player 2. At this time, players must choose either the left or right arena by moving their avatar in the direction of their preferred arena. Once an avatar contacts an arena, the joystick control for that avatar is temporarily relinquished. After both players choose their preferred arena, another 750-ms interphase-interval delay is imposed, and then the task moves to the second phase. In phase 2, players are randomly assigned to be the actor and observer, the red rectangles disappear, and the two avatars are displaced from the edge to the bottom interior of the corresponding arenas. The actor is placed 10.1 cm away from the vertical midline and 6 cm below the horizontal midline. The observer is placed 10.1 cm away from the vertical midline and 12 cm below the horizontal midline. Immediately afterwards, the actor starts playing the token-capture game on its monitor, a copy of which is shown on the observer's monitor. The observer was free to move the joystick, but joystick movements were disconnected from avatar control during observer trials. The actor must use the joystick to move the avatar left or right to capture 15 falling tokens (grey circular disks of diameter 6.72 cm). The tokens drop sequentially every 1/3 s, starting 16.8 cm above the centre and moving directly downward at the constant speed of 50.4 cm s^{-1}. The horizontal positions of tokens are sampled from a Gaussian process centred at the initial position of the actor's avatar with a squared exponential kernel. To vary task difficulty across trials, the s.d. of the Gaussian process was sampled from a discrete uniform distribution ([2.52, 5.04, 7.56] cm).

Each trial ends in either a win or a lose state. To maximize wins, the actor must select the correct arena and collect as many tokens as possible. The correct arena (the one associated with a non-zero win probability) switches covertly in a blocked fashion (see below). In either arena, captured tokens turn green, whereas missed tokens disappear without effect. A win is signalled by a change of the colour of the avatars and an auditory tone. A trial ends either with a win or after all 15 tokens pass. The observer receives no reward but can monitor trial progression and infer the outcome from the visual and auditory feedback. Each trial lasts 5.42 ± 2.18 s. After the trial, the display remains stationary for 1 s before switching to a uniform black screen for an inter-trial interval of 2.676 ± 0.085 s.

Before the two-player game, players completed a single-player version with a single avatar, designated as the actor. All other task aspects remained identical to the two-player version. The task contingencies were identical for humans and monkeys with a few exceptions as follows.

Non-human primates. (1) The switching probability was 0 for the first 10 trials, 1/3 for trials 11–24 and 1 for trial 25. (2) In the correct arena, captured tokens may trigger a win with a 15% probability. (3) A win triggered a juice drop for the actor. (4) Animals first played the single-player version until their performance stabilized. The single-player data reported are from sessions with stable performance. (5) They were then introduced to the two-player version without electrophysiology recordings until performance stabilized. Reported behavioural and neural data are from subsequent sessions with simultaneous behavioural and physiological recordings.

Humans. (1) The switching probability was 0 for the first 10 trials and 1/3 for subsequent trials. (2) In the correct arena, captured tokens may

trigger a win with a 10% probability. (3) Before the start of the first session, participants received written instructions about the game but were not informed of the reward probabilities or state-switching statistics. (4) Participants played the single-player version for five or six sessions (see below). Single-player sessions started with 50 practice trials where the correct and incorrect arenas were cued (green and yellow, respectively). After an optional short break, participants completed 600 trials of the one-player game with 2-min breaks every 200 trials. Reported single-player data are from sessions 1–5 and do not include the practice trials. (5) Initially, participants were scheduled for five single-player sessions. During data collection, a confidence-reporting step was added, requiring participants to use a keyboard to report their confidence on every trial (1, 'not confident at all', to 4, 'fully confident') after selecting the arena, in both single- and two-player games. Nine participants who completed five sessions of single-player in the original task (that is, with no confidence report) played a sixth session with confidence reporting, and the tenth participant used the augmented task from the start and did not require a sixth session. (6) Participants received instructions about the two-player game before the first two-player session. Each session included 525 trials with two 2-minute breaks after every 175 trials. Each pair completed 15 two-player sessions, totalling 7,875 trials.

Analysis of behaviour

We used players' choice behaviour to quantify the probability of choosing the correct arena from four trials before to ten trials after a block switch (Fig. 1h–k).

Solipsistic agent. To test whether a player was sensitive to observer trials, we compared its performance to the performance predicted if the player were to ignore all observer trials and treat the two-player game as single-player. We refer to this hypothetical player as solipsistic. If we denote a win state at trial n by R_n, then the average probability of R_n for a solipsistic actor in the two-player game (P_{2p}) on the basis of their average performance in the one-player game (P_{1p}) can be written as follows:

$$\langle P_{2p}(R_n)\rangle = \sum_{0}^{n} \binom{n}{k}\left(\frac{1}{2}\right)^n \langle P_{1p}(R_{n-k})\rangle \quad (1)$$

The <.> denotes average probabilities. The sum runs over the observer trials preceding trial n since the last block switch, indexed by k. To understand this equation, consider a sequence of n trials with k observer and $n-k$ actor trials. The binomial coefficient (n-choose-k) counts such combinations, whereas $(½)^n$ gives their probability, with $(½)^k$ for k observer and $(½)^{n-k}$ for $n-k$ actor trials. The term $P_{1p}(R_{n-k})$ implements the assumption that the actor disregards observer trials, behaving as if only $n-k$ trials have passed since the last block switch.

We tested the significance of the difference in accuracy between this agent and the participant using t-tests on the average proportion correct (P(correct)) for positions [0,10].

Oracle agent. To estimate an upper performance bound on trial n, we simulated an oracle agent who mimicked the player's choices for trial 1:$n-1$ but selected the correct choice on trial n. Therefore, if the correct choice on trial n is to switch, regardless of whether trial $n-1$ was rewarded or not, the oracle will switch. This results in a non-zero probability of switching after rewarded trials.

Difficulty of collecting tokens

In the second phase, token positions were random, creating varying difficulty levels. To assess whether players attempted to maximize captured tokens, we analysed performance on unrewarded trials as a function of difficulty (D), defined as the sum of absolute distances between successive tokens:

$$D = \sum_{0}^{14} |x_{i+1} - x_i| \quad (2)$$

Here, x_i is the horizontal position of the ith token in screen coordinates. Because each unrewarded trial has 15 tokens, the difficulty is the sum of 14 horizontal displacements.

Subjective belief about the correct arena and block switches

Because the correct arena and block switches were covert, we developed a method to estimate participants' subjective beliefs about the arena and trial position within a block. The first rewarded trial of a session was assumed to indicate the correct arena and was assigned position 1 of the first block. The position incremented until the first reward on the opposite arena, which was assumed to signal a block switch, resetting the position to 1. Early, mid and late trials were defined as [1,5], [6,10] and [11,∞), with bin sizes balanced as closely as possible.

Regression analysis of switch behaviour and confidence

We used logistic regression to quantify the dependence of switch behaviour on various factors after unrewarded trials, including the number of consecutive unrewarded trials, the number of tokens captured and position in the trial:

$$\text{Switch} = \beta_u N_{\text{unrewarded}} + \beta_t N_{\text{tokens}} + \beta_p N_{\text{position}} + \beta_0 \quad (3)$$

Switch is a binary variable indicating when the player chooses a different side than the actor's current choice (that is, the arena for which direct evidence is acquired). $N_{\text{unrewarded}}$ is the number of consecutive unrewarded trials in the same arena. We included only up to four consecutive unrewarded trials in this analysis. N_{tokens} is the number of touched tokens and N_{position} is the subjective trial position in the block. β are fitted parameters controlling the influence of each variable on switch behaviour.

Combining both actor and observer trials, we used simple logistic regression (for each monkey) or mixed-effects logistic regression (for human participants):

$$\text{Switch} = \beta_c I_{\text{condition}} + \beta_u N_{\text{unrewarded}} + \beta_t N_{\text{tokens}} + \beta_p N_{\text{position}} + \beta_0 \quad (4)$$

$I_{\text{condition}}$ is a binary indicator variable (0 for actor, 1 for observer).

We also computed the contribution from choice conflict for the actor condition. For this analysis, we included incongruent trials, which were excluded in previous regressions:

$$\text{Switch} = \beta_u N_{\text{unrewarded}} + \beta_t N_{\text{tokens}} + \beta_p N_{\text{position}} + \beta_{cg} I_{\text{congruence}} + \beta_0 \quad (5)$$

$I_{\text{congruence}}$ is a binary indicator variable (1 for incongruent, 0 for congruent).

For human participants, we additionally performed mixed-effects linear regression on their confidence report:

$$\text{Confidence} = \beta_c I_{\text{condition}} + \beta_u N_{\text{unrewarded}} + \beta_t N_{\text{tokens}} + \beta_p N_{\text{position}} + \beta_0 \quad (6)$$

Analysis of single neurons

We estimated each neuron's firing rate by averaging spike counts in shifting 100-ms time bins with a 10-ms step size. We analysed firing rates aligned to different task events including the choice time (that is, when the avatar contacts an arena) and outcome (that is, reward time in rewarded trials and end of token collection in unrewarded trials). For visualization, binned firing rates were smoothed using a three-bin moving average.

Single-neuron sensitivity to outcome and choice

We measured selectivity to reward outcome in self/other conditions using receiver operating characteristic (ROC) analysis on the basis

of spike counts within 600-ms windows, either after the outcome or before the choice. The ROC score, calculated as the area under the performance curve of a binary classifier, classified trials as rewarded or unrewarded—for either the current trial (outcome) or the previous trial (choice). To centre selectivity at 0, we subtracted 0.5 from the score, where values of −0.5 and 0.5 indicate perfect separation of reward outcome with lower or higher firing rates for the reward condition, respectively. Significance was assessed using bootstrap (1,000 iterations, $P < 0.05$). To compute the correlation of ROC scores between self and other conditions, we performed total least squares regression between self and other selectivities.

Single-neuron sensitivity to cumulative errors

We computed selectivity for the history of congruent unrewarded trials (1NR versus 2NR) using ROC analysis for neurons that exhibited significant reward selectivity in either the actor or observer condition and maintained the same selectivity sign in both. Binary classifiers were constructed for the four possible consecutive outcomes: 1NR–actor versus 2NR–actor (AA), 1NR–actor versus 2NR–observer (AO), 1NR–observer versus 2NR–actor (OA) and 1NR–observer versus 2NR–observer (OO). For neurons with significant selectivity in any condition, we also calculated the average differences in *Z*-scored firing rates across rewarded actor trials, 1NR (actor or observer) and 2NR (actor or observer). Neurons with consistent rate changes (either 1R < 1NR < 2NR or 1R > 1NR > 2NR) were classified as encoding cumulative error.

Single-neuron sensitivity to actor and observer conditions

We compared trial-by-trial firing rates in the 600 ms after outcome between actor and observer conditions using a rank-sum test, separately for rewarded and unrewarded trials. For neurons with significant differences, we calculated the absolute *Z*-scored firing-rate differences between actor and observer conditions.

Population activity for actor and observer outcome and SE

We used targeted dimensionality reduction (TDR) to identify encoding dimensions of actor outcome, observer outcome and SE[47]. To do so, we used regression to relate each neuron's spike count within a 600-ms window following the outcome to different task variables.

For actor and observer outcome (analysed separately), we used the following regression:

$$Z = \beta_{\text{outcome}} I_{\text{R/N}} + \beta_{\text{choice}} I_{\text{L/R}} + \beta_0 \quad (7)$$

$I_{\text{R/N}}$ is an indicator variable for outcome (1 for unrewarded, 0 for rewarded), and $I_{\text{L/R}}$ an indicator variable for choice (−1 for left, 1 for right). To reduce estimation noise, only neurons that had more than five trials in all conditions were included.

For actor and observer outcome without rewarded trials, we used the following regression:

$$Z = \beta_{\text{history}} I_{\text{1NR/2NR}} + \beta_{\text{choice}} I_{\text{L/R}} + \beta_0 \quad (8)$$

$I_{\text{1NR/2NR}}$ is an indicator variable for history (0 for 1NR trials, 1 for 2NR trials).

For SE, we used the following regression:

$$Z = \beta_{\text{switch}} N_{\text{preswitch}} + \beta_{\text{choice}} I_{\text{L/R}} + \beta_0 \quad (9)$$

$N_{\text{preswitch}}$ represents the distance in trials from the next switch, with values of −1 for one trial before, −2 for two trials before and −3 for more than two trials before the switch. $I_{\text{L/R}}$ is an indicator variable for choice (−1 for left, 1 for right). To reduce estimation noise, only neurons that had more than ten trials in all conditions were included.

After solving the regression for all neurons, we arranged the coefficients in a matrix and orthogonalized columns using *QR*-decomposition, requiring *R* to have positive diagonal values such that the columns of *Q* provided are orthogonalized set of coefficients for each variable.

For actor and observer (equation (7)), we used the orthogonalized coefficients associated with the outcome (first column) as the actor and observer outcome dimension for each condition. For SE (equation (9)), we used the coefficients associated with $N_{\text{preswitch}}$ (first column) as the SE dimension. For cross-validation of the SE dimension, we randomly selected one trial per condition, computed the evidence dimension from the remaining trials and projected the held-out trial activity onto this dimension. This process was repeated 100 times, generating 100 cross-validated projections per condition.

Predicting switch behaviour from SE dimension

We projected neural activity onto the SE dimension to predict switch behaviour in the next trial. For each recording session, we selected trials with at least two recorded neurons and at least ten switch trials. Using spike counts from all but one randomly selected trial, we computed the SE dimension and projected activity from all trials onto it, generating a distribution of projection values. The held-out trial was classified as high or low evidence on the basis of whether its projection value was above or below the median of this distribution. This process was repeated 1,000 times per session. We then calculated the proportion of switch trials in the high- and low-evidence groups, considering sessions significant if the low-evidence group had a significantly lower proportion of switches, as determined by a rank-sum test.

Contribution of actor and observer outcome to SE dimension

We computed neural SE in actor and observer conditions by randomly selecting one trial from each actor and observer × 1R/1NR/2NR condition as a held-out test trial, deriving the evidence dimension from the remaining trials and projecting the test trial activity onto this dimension. Only accumulation-selective neurons from sessions where neural SE significantly predicted switching behaviour were included. This process was repeated 100 times to generate a distribution of projection values for each condition.

The geometry of actor and observer outcome and SE

We measured pairwise angles between actor outcome, observer outcome and SE dimensions using a randomly selected half of the trials and repeated this process 100 times, generating a distribution of angles. We also measured the angle between actor and observer separately in two subspaces: the aligned subspace, computed from those neurons whose coefficients in actor– and observer–outcome coefficients had the same sign; and the anti-aligned subspace, which had opposite signs.

Stability of SE in the inter-trial interval

We computed the SE dimension using a sliding window of 600 ms with a 300-ms step size over the 3 s following outcome onset. For each time window, we calculated the angle between the SE dimension at that time and the SE dimension computed immediately after the outcome.

Analysis of neural geometry and behaviour across sessions

We measured pairwise neural angles between outcome and SE dimensions using data from individual sessions. To test whether these angles were predictive of learning rate, we performed a regression analysis relating neural angle to behavioural sensitivity, defined as the slope of the regression between *P*(switch) and the number of unrewarded trials. We carried out this analysis across several data subsets: (1) combined across animals and conditions (actor and observer); (2) combined across animals but separately for actor and observer conditions; (3) separately for each animal, combining across conditions; (4) separately for each animal and each condition.

Because sessions varied in the number of recorded neurons, we used weighted total least squares regression, assigning weights on the basis

of the expected variance in angle estimation. Specifically, we estimated each session's weight using the following procedure.

We first identified the session with the maximum number of neurons (N) and computed its outcome–switch angle (θ). This session was used as the reference and assigned a weight of 1. To estimate the reliability of angle measurements in sessions with fewer neurons ($M < N$), we simulated how angle estimates degrade with reduced dimensionality. We constructed two vectors in N-dimensional space with a known angular separation θ and then repeatedly (5,000 times) subsampled M random dimensions and computed the angle between the truncated vectors. This yielded a distribution of estimated angles for dimensionality M. The variance of this distribution reflects the expected noise in angle estimation for a session with M neurons. We then used the inverse of this variance as the weight assigned to that session in the regression.

Session-level analysis of neural geometry and behaviour

We measured pairwise angles using data from single sessions. To assess the correlation between a neural angle (between outcome and SE dimensions) and rate of learning (the regression slope of P(switch) over number of unrewarded trials), we performed weighted total least squares regression between the neural angle and the behavioural slope for actor and observer conditions. The weight of each session was determined by the expected variance of measurement given the number of neurons available in that session, relative to the maximum number of neurons recorded in any session. Given the session with the maximum number of neurons N and measured angle θ, we constructed two vectors **V1** and **V2** in N-dimensional space with θ between them. For this session, the weight was set to 1. For each session with number of neurons $M < N$, we subsampled **V1** and **V2** in randomly selected M dimensions from the original N-dimensional space and computed the angle between them. We repeated this process 5,000 times to obtain the s.d.(M). The inverse of this variance was used as the weight in the regression.

To assess the degree to which either animal's learning rate as an observer was correlated with the actor's, we performed total least squares regression between the behavioural slope in actor condition for M1 (M2) and observer condition for M2 (M1).

Neural network model for multi-agent integration task

We used RNN models with different architectural and optimization constraints to test two hypotheses about how the ACC integrates experiential and observational evidence into cumulative switch belief. One hypothesis (H1) posits that the ACC receives the experienced and observed evidence through a common identity-agnostic input while another input provides information about the identity (self/other). The other hypothesis (H2) posits that the ACC receives independent inputs for experienced and observed outcomes.

Architectural constraints. All models have three layers: an input layer providing three distinct inputs, a hidden layer consisting of 200 recurrently connected units and an output layer for computing the network output.

RNNs instantiating H1 receive one input conveying information about both experienced and observed outcomes and another input for identity (actor, −1; observer, 1). RNNs instantiating H2 receive experienced and observational outcomes through separate inputs, with only one of them active in any given trial (the other is set to 0). These inputs project to all hidden units, are active at the onset of each trial and are set to 0 at other timesteps. In both cases, a third input serves as a go cue instructing when the RNN has to generate an output, Y, which is a scalar reflecting the cumulative evidence across trials.

For all RNNs, the input carrying outcome information is a sample from a bimodal Gaussian distribution (equation (10)), with one positive and one negative mode centred at 0.5 and −0.5, corresponding to win (rewarded) and lose (unrewarded) states, respectively:

$$P(x) = \frac{1}{2} \times N\,(x|\mu = -0.5, \sigma = 0.1) + \frac{1}{2} \times N\,(x|\mu = 0.5, \sigma = 0.1) \quad (10)$$

Here, μ is the mean and σ is the s.d. For both H1 and H2, we tested two model variants. In one variant, we assumed the readout weights that drive the output were learnable ($H1^{learn}$, $H2^{learn}$), and in the other, the readout weights were initialized randomly and were not learnable ($H1^{fix}$, $H2^{fix}$). The weights of the input layer were initialized randomly and were not subjected to learning.

Optimization constraints. All RNNs were trained to adjust their output, Y, according to the following requirements:

(1) Y must reset to 0 after rewarded trials (positive inputs).
(2) Y must integrate outcome input following unrewarded trials (negative inputs) for both actor and observer conditions and maintain the integrated value across trials.
(3) To replicate the actor and observer behavioural asymmetry, actor and observer inputs must be integrated with a gain of 1 and 0.5, respectively.
(4) In the first variant ($H1^{learn}$, $H2^{learn}$), both the recurrent and readout weights were trained. In the second variant ($H1^{fix}$, $H2^{fix}$), training was applied only to the recurrent weights.
(5) In the third variant ($H1^{wide}$, $H2^{wide}$), we explicitly encouraged separation between actor and observer outcome representations by adding a penalty term to the loss function proportional to the squared cosine similarity between the corresponding readout vectors. In principle, networks under H1 may learn effectively orthogonal representations by forming two subpopulations, with each representing actor and observer outcome using the identity input as a gate. In practice, however, this solution may be too fragile and difficult to reach through gradient descent.

Information was presented in a trial-based manner, with each trial having a variable duration sampled uniformly between 10 to 20 timesteps. The input was provided at the first timestep and then set to 0. A go cue input instructed the RNN when to generate an output. This go cue was presented as a linear ramp from 0 to 1 over five timesteps, remaining at 1 for a further five timesteps. The onset of the go cue, relative to the trial start, was sampled uniformly from zero to five timesteps. RNNs were required to compute and maintain the output while the go cue was 1.

Model dynamics. The activity of hidden units is given by

$$\tau \frac{dx}{dt} = -x(t) + W_e I(t) + W_i f(x(t)) + b \quad (11)$$

$$Y(t) = W_o x(t) \quad (12)$$

In equation 11), $\tau = 5$ is the time constant, t is the timestep, $x(t)$ is the activity of all units, W_e is the embedding weights for inputs $I(t)$, W_i is the recurrent weights, f is a tanh nonlinear function and b is a bias term. In equation (12), $Y(t)$ is the output and W_o is the readout weights. All parameters were randomly initialized from a normal distribution with zero mean and variance $1/N$, where N is the number of parameters for each layer. We trained the networks using gradient descent in batches of 16, by minimizing the mean squared error loss between output and target (when the go cue is at level 1). We do not constrain network dynamics outside of the reporting window.

We trained 100 models with random initiations per hypothesis, for a total of 600 models, and each model was trained for 100,000 iterations, with 500 timesteps per iteration.

Model performance. We computed the performance of trained networks as the average output in the reporting period conditioned on input and trial history. 1R, 1NR and 2NR were assigned the same way as in behavioural trials.

Model analysis. Similar to the ACC, we applied targeted dimensionality reduction to activity in the hidden layer units to identify encoding dimensions for actor outcome, observer outcome and SE.

For actor and observer outcome (analysed separately), we used the following regression:

$$Z = \beta_{\text{input}} V_{\text{input}} + \beta_0 \quad (13)$$

V_{input} is an indicator variable for outcome (1, unrewarded; 0, rewarded).

For SE, we used the following regression:

$$Z = \beta_{\text{output}} V_{\text{output}} + \beta_0 \quad (14)$$

V_{output} is the output value at the time when the go cue reaches 1 for each trial.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data are available via DANDI at https://dandiarchive.org/dandiset/001435. Source data are provided with this paper.

Code availability

The code for analysis in this paper is available via GitHub at https://github.com/jazlab/RC_SR_NV_SBY_MJ_2025.

65. Watters, N., Tenenbaum, J. & Jazayeri, M. Modular object-oriented games: a task framework for reinforcement learning, psychology, and neuroscience. Preprint at https://arxiv.org/abs/2102.12616 (2021).

Acknowledgements R.C. was supported by the Simons Center for the Social Brain at MIT and Hock E. Tan and K. Lisa Yang Center for Autism Research. S.R. was supported by a Mathworks Graduate Fellowship and K. Lisa Yang ICoN Center Fellowship. S.B.M.Y. was supported by the Simons Center for the Social Brain at MIT. M.J. was supported by the Simons Foundation and the McGovern Institute. We thank J. Gabel, N. Watters and A. Piccato for their respective help with electrophysiology, modelling and open-sourcing.

Author contributions S.R., S.B.M.Y. and M.J. designed the task. S.B.M.Y. trained the animals. R.C. and N.V. performed animal experiments. S.R. performed human experiments. R.C., S.R. and N.V. analysed the data. R.C., S.R. and M.J. wrote the manuscript. M.J. supervised the project.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09885-0.
Correspondence and requests for materials should be addressed to Mehrdad Jazayeri.
Peer review information *Nature* thanks Kayvon Daie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

nature portfolio

Corresponding author(s): Mehrdad Jazayeri

Last updated by author(s): Oct 2, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Open Ephys version 0.5.5.4 was used to collect neural data. Kilosort version 3.0 was used for spike sorting. Phy version v2 .0bl was used to verify spike sorting quality manually. Mworks version 0.11 and MOOG version 1.4 was used to collect behavioral data.
Data analysis	Custom python code (python version 3.9) was used to analyze data. Code is available at github.com/jazlab/RC_SR_NV_SBY_MJ_2025

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data is available on DANDI at https://dandiarchive.org/dandiset/001435.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Self-reported gender of participants were: 6 males, 4 females. Gender-based analyses were not performed because the study was not designed to ask gender-specific questions.
Reporting on race, ethnicity, or other socially relevant groupings	Race, ethnicity data was not collected.
Population characteristics	Participants reported age range was 18-65.
Recruitment	Participants were recruited by email to potential participants near MIT. This could bias participants to students and researchers. We do not have expectation of any impact on our results.
Ethics oversight	The study was approved by the MIT Committee on the Use of Humans as Experimental Subjects

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We used n= 2 monkeys, with n=36 sessions. Number of sessions was sufficient to replicate findings by cross validation across sessions.
Data exclusions	We did not exclude data relevant to this study.
Replication	We used cross-validation to check robustness of our results. In cases where data was pooled from subjects, sessions, or conditions, we also performed analysis on individual subjects, sessions, or conditions, and report results from each. The behavior experiment was repeated for 36 sessions, with independent stimuli sequences. Replication was successful as determined by analysis of individual sessions.
Randomization	Trial conditions were presented randomly.
Blinding	Trial conditions were randomly sampled and therefore both the experimenters and subjects were blind to the experimental conditions.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	2 rhesus macaque monkeys (Macaca mulatta), aged 6 and 11 years old.
Wild animals	No wild animals were used.
Reporting on sex	The monkeys were 1 female and 1 male.
Field-collected samples	No field-collected samples were used.
Ethics oversight	All procedures were performed in compliance with the guideline of National Institutes of Health and American Physiological Society, and approved by the MIT Committee on Animal Care.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*