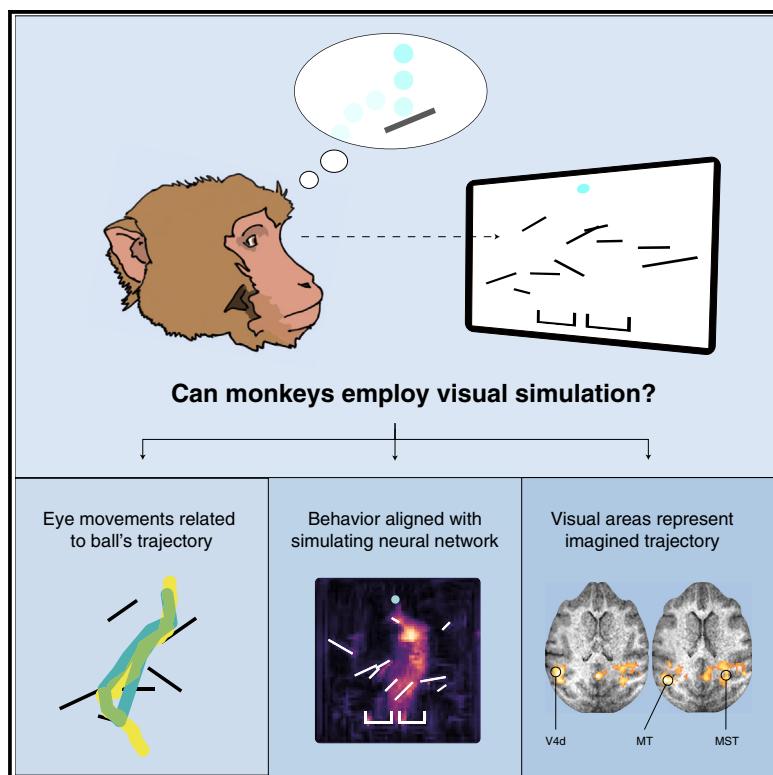


Monkeys engage in visual simulation to solve complex problems

Graphical abstract



Authors

Aarit Ahuja, Nadira Yusif Rodriguez,
Alekh Karkada Ashok, Thomas Serre,
Theresa M. Desrochers,
David L. Sheinberg

Correspondence

david_sheinberg@brown.edu

In brief

Ahuja et al. show that, like humans, monkeys can employ visual simulation to predict future motion trajectories. Functional MRI further reveals that motion-sensitive brain regions likely represent the imagined trajectories during this process, suggesting that such simulations may evoke perceptible sensory representations.

Highlights

- Monkeys employed simulation to predict the path of a ball while playing “Planko”
- Monkeys’ strategies matched RNNs that learned a simulation-based task approach
- fMRI showed motion-sensitive brain activity during visual simulation in monkeys
- These results suggest that primates can perform imaginative visual simulations



Article

Monkeys engage in visual simulation to solve complex problems

Aarit Ahuja,^{1,2} Nadira Yusif Rodriguez,¹ Alekh Karkada Ashok,³ Thomas Serre,^{3,4} Theresa M. Desrochers,^{1,4,5} and David L. Sheinberg^{1,4,6,*}

¹Department of Neuroscience, Brown University, Meeting Street, Providence, RI 02906, USA

²Exponent, Worcester Street, Natick, MA 01760, USA

³Department of Cognitive and Psychological Science, Brown University, Thayer Street, Providence, RI 02906, USA

⁴Robert J. and Nancy D. Carney Institute for Brain Sciences, Brown University, Angell Street, Providence, RI 02906, USA

⁵Department of Psychiatry and Human Behavior, Brown University, Providence, RI 02906, USA

⁶Lead contact

*Correspondence: david_sheinberg@brown.edu

<https://doi.org/10.1016/j.cub.2024.10.026>

SUMMARY

Visual simulation—i.e., using internal reconstructions of the world to experience potential future versions of events that are not currently happening—is among the most sophisticated capacities of the human mind. But is this ability in fact uniquely human? To answer this question, we tested monkeys on a series of experiments involving the “Planko” game, which we have previously used to evoke visual simulation in human participants. We found that monkeys were able to successfully play the game using a simulation strategy, predicting the trajectory of a ball through a field of planks while demonstrating a level of accuracy and behavioral signatures comparable with those of humans. Computational analyses further revealed that the monkeys’ strategy while playing Planko aligned with a recurrent neural network (RNN) that approached the task using a spontaneously learned simulation strategy. Finally, we carried out awake functional magnetic resonance imaging while monkeys played Planko. We found activity in motion-sensitive regions of the monkey brain during hypothesized simulation periods, even without any perceived visual motion cues. This neural result closely mirrors previous findings from human research, suggesting a shared mechanism of visual simulation across species. Taken together, these findings challenge traditional views of animal cognition, proposing that nonhuman primates possess a complex cognitive landscape, capable of invoking imaginative and predictive mental experiences to solve complex everyday problems.

INTRODUCTION

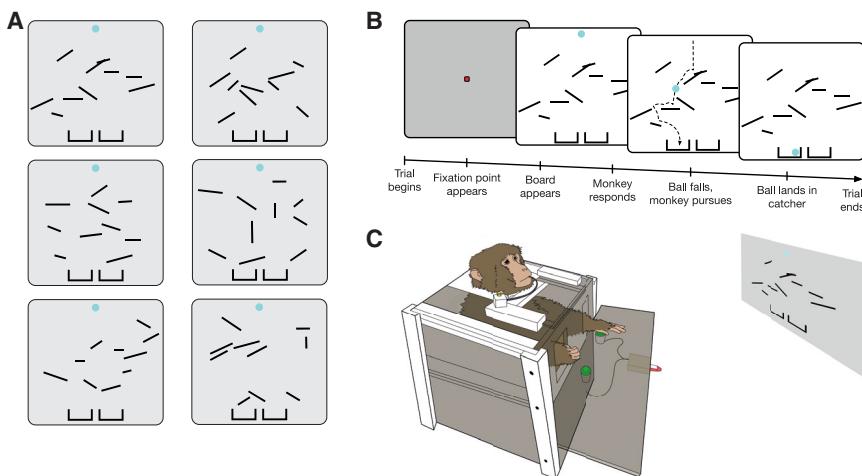
Consider the following scenario: it’s a Monday morning, and you’re driving to work. You’ve taken this route dozens of times and can navigate without a second thought. Yet, today, as you approach your destination, you encounter construction blocking your usual path. You stop and start to think, “Maybe I could turn right here and head north on a parallel road? Although that does lead to a one-way street. Perhaps I’ll have better luck if I take a left? That should take me past the bakery, around the school crossing, and eventually direct me back to my destination.” In just a few seconds, you have managed to simulate potential alternate paths and chart a new course for your journey.

This process of problem solving via “mental simulation” takes place entirely in your head, without the need to move a single muscle. You can imagine how things might play out in the future to help you arrive at a solution. Importantly, you need not actually experience the things you are simulating (or their potentially negative consequences, such as going down a one-way street). Your internally generated mental recreations are sufficient to guide your actions. When harnessed effectively, mental simulation is one of the

most sophisticated and useful cognitive capacities at your disposal.

Mental simulations can take various forms. For instance, past research has shown that imagining bodily movements relies on a form of mental simulation, as indicated by the overlap in neural circuits involved in imagination and execution of motor actions. This phenomenon is referred to as “action simulation.”¹ Similarly, mental simulation strategies can also explain how we predict outcomes in physical scenes, such as the collapse of a Jenga tower.^{2,3} In recent experiments, we have built upon previous intuitive physics studies to provide evidence for a specific type of mental simulation called visual simulation.^{4,5} As the name implies, visual simulation involves imagined visual representations corresponding to the objects involved in the mental simulation, similar to how action simulation involves imagined motor representations. In other words, visual simulation incorporates a distinctly visual component into the simulation process. Supporting this idea, we have developed a novel task called Planko (Figure 1) and demonstrated that when people are asked to predict the likely trajectory of a ball falling through an obstacle-filled display, their behaviors and eye movements indicate that they are simulating the ball’s path.⁵ Furthermore, we have shown



**Figure 1. Experimental setup**

(A) Examples of Planko boards used in the task. Monkeys were required to predict which catcher the ball would land in if dropped. In the present example, the three boards in the left column lead to the left catcher, and the three boards in the right column lead to the right catcher.

(B) A schematic of one complete trial, including the pre-response period when the monkeys could potentially simulate the ball's trajectory and the post-response period when they saw the ball fall.

(C) A diagram of the NHP upright rig setup that was used for training and behavioral testing on the task. Monkeys indicated their responses using one of the two provided buttons and were given juice reward for correct responses.

that during these simulations of the ball's trajectory, motion-sensitive brain regions like the middle temporal area (MT) respond as if the ball's motion were being literally seen, even though the stimulus remains static throughout the simulation.⁴ Collectively, these findings suggest that humans are indeed capable of visual simulation, and the neural correlates of this process can be observed in visual brain areas.

Despite the growing body of behavioral and neuroimaging work on simulations in the brain, several questions about the underlying neural mechanisms that support these phenomena remain. Why is this? We suggest that a major obstacle to progress derives from the lack of a compelling animal model. The absence of research on mental and visual simulation in animals is not surprising, given the complexity, introspection, and subjectivity associated with these phenomena. Although some recent evidence suggests that computational models of simulation align with nonhuman primate (NHP) behavior, it remains unclear whether animals are capable of mental simulation, let alone visual simulation.⁶ In our current experiments, we aimed to address these questions by replicating our human studies on visual simulation with NHPs. We found that when macaques play Planko, their behavioral patterns can be accurately accounted for by models assuming a simulation strategy. Using awake monkey fMRI, we further discovered that when monkeys engage in a simulation of the ball's trajectory, motion-sensitive brain regions become active, indicating an explicitly visual aspect of the simulation process. Previously, we provided evidence for these same findings with human participants. Together, these results demonstrate that monkeys not only possess the ability for visual simulation but also share the biological foundations of this capability with humans and other NHPs.

RESULTS

Can monkeys play Planko?

We have developed a task paradigm called Planko to probe visual simulation. During the task, participants are shown displays (also referred to as boards), like the ones shown in Figure 1A, and asked to predict in which of the two bottom "catchers" the ball would end up were it to be dropped through a field of randomly arranged "planks." While the participants make this prediction

(i.e., during the "pre-response period"), the ball remains completely static, suspended in place. Once the participants indicate their choice with a button press, the ball is then in fact dropped during the "post-response period," thus providing participants with feedback on their responses. Although we have been successful in using this task to probe simulation in humans,⁵ we wondered whether monkeys would be capable of learning it and, if so, would they, too, rely on visual simulation?

We thus set out to train two monkeys (referred to here as monkey G and monkey A) to play Planko. We started with extremely simple displays, containing very few planks (Figure 2B). In the early stages, we also introduced a "shadow ball," which left a light gray trace of the ball's trajectory on random trials (Figure 2A), showing the correct answer and helping the monkeys focus on the ball's movement. Over time, we increased the number of planks while reducing the use of the shadow ball and gradually fading its trajectory (Figures 2A and 2B). We assessed training success by analyzing monkeys' accuracy on non-shadow trials.

The progression of each monkey's accuracy is shown in Figure 2C. Initially, both monkeys struggled when more planks were introduced, especially in the first 25 sessions. For example, monkey G quickly learned the one-plank version but reverted to chance-level accuracy when two planks were added. Monkey A showed a similar decline in performance as plank numbers increased. This suggests their early strategies were not robust to visual changes and weren't based on simulation. However, by session 25, both monkeys improved significantly, achieving accuracy regardless of plank count, even when encountering new displays. This newfound invariance to the visual properties of the scene is striking and suggests that the monkeys were able to arrive at a more sophisticated strategy (like simulation) for solving the task (Video S1). To further confirm this notion, we tested both monkeys' behavior against several "shortcut" rules that could produce over 50% accuracy without simulating the ball's trajectory. The monkeys' behavior matched none of these rules (Figure S1), reinforcing the idea that they were employing a highly sophisticated strategy for solving the task.

Behavioral evidence for simulation in monkeys

Although we were encouraged by the fact that the monkeys were able to play Planko with higher-than-chance accuracy, that by

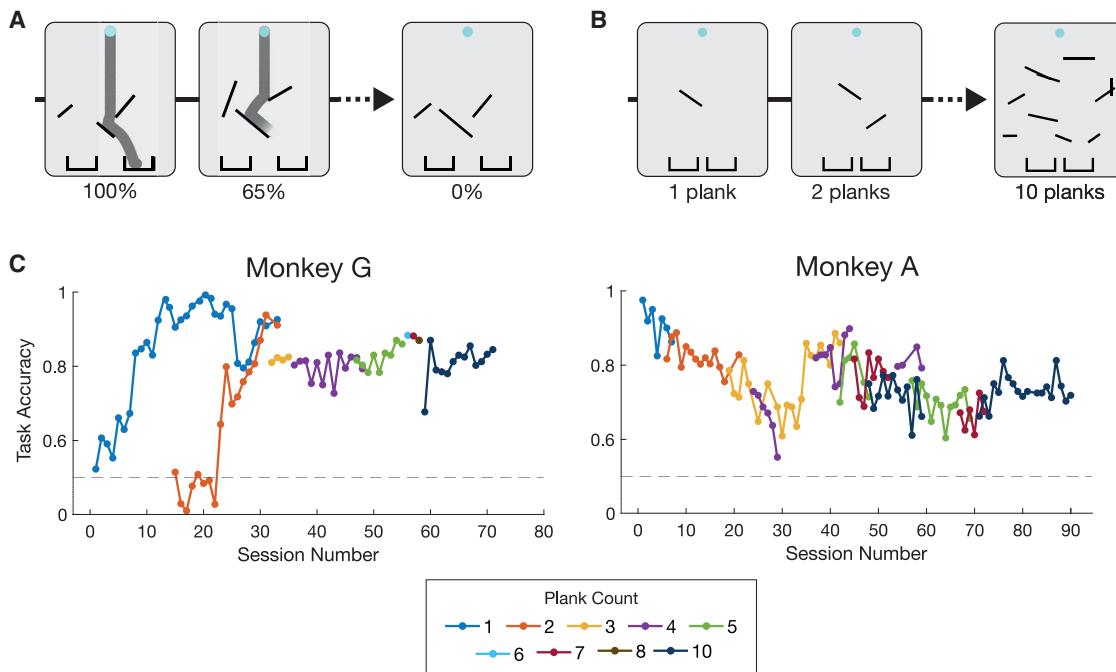


Figure 2. Training monkeys to play Planko

(A) Examples of the “shadow ball” used to train the monkeys on Planko. Shadow ball trials were mixed with non-shadow ball trials and gradually faded to reveal none of the trajectory, encouraging monkeys to extrapolate the ball’s path.

(B) Examples of varying plank counts used during training. Both monkeys began with one plank and progressed to navigating ten planks.

(C) Task accuracy progression across sessions as plank numbers increased (non-shadow ball trials only). Both monkeys initially struggled but later maintained consistent accuracy, indicating a sophisticated, generalizable strategy.

See also [Figure S1](#) and [Video S1](#).

itself does not mean that they were doing so by using simulation. We thus set out to ascertain whether their behavior on the task was in line with what might be expected were they engaging in visual simulation. In our previous work, we have shown that human participants’ accuracy on the task depends on the degree of “simulation uncertainty” created by the plank configuration on any board (for details on simulation uncertainty, see [STAR Methods](#), [Figures 3A, 3B](#), and [S2](#); [Video S2](#)). In this study, monkeys were shown boards that fell into one of two discrete simulation uncertainty categories—low or high ([Figure 3C](#)). We predicted that if monkeys were engaging in visual simulation, then their task accuracy would decrease as simulation uncertainty increased. [Figure 3D](#) shows both monkeys’ task performance as a function of simulation uncertainty (high vs. low). As with humans ([Figure 3G](#)), we found that both monkeys were significantly worse at the task when simulation uncertainty increased (monkey G: $t_{190} = 2.75$, $p < 0.005$; monkey A: $t_{190} = 3.27$, $p < 0.005$), suggesting that they were employing a simulation strategy to approach the task.

We also compared monkeys’ eye movements before and after their response on each trial. In the pre-response period, they made saccades while looking at the static image of the board. In the post-response period, they followed the falling ball with smooth-pursuit eye movements. Our goal was to determine whether the eye movements made while trying to determine the ball’s final position significantly overlapped with the eye movements made while perceiving the ball’s actual falling

trajectory (see [STAR Methods](#), [Figures 3E](#) and [S2](#)). We predicted that if monkeys were visually simulating the ball’s movement path to solve the task, then there would be significant overlap between pre-response and post-response eye movements. [Figure 3F](#) shows both monkeys’ overlap between pre- and post-response eye movements relative to overlap predicted by chance (see [STAR Methods](#)). For both monkeys, we found that the degree of observed spatial overlap was greater than chance (monkey G: $t_{239} = 3.49$, $p < 0.001$; monkey A: $t_{240} = 4.17$, $p < 0.001$), suggesting that they may have relied on visual simulation to arrive at the correct answer. This result also precisely mirrors what we have previously observed with human participants ([Figure 3G](#)).

Computational evidence for simulation in monkeys

To develop a computational explanation for playing Planko using simulation and non-simulation strategies, we trained two neural networks: a shallow convolutional neural network (CNN) and a recurrent neural network (RNN) ([Figure 4A](#)). We opted for this approach because in previous work we have shown that such networks (especially RNNs) are capable of spontaneously developing task strategies that resemble visual simulation.⁷ The CNN architecture we used in this study was based on a previous model used in our human research,⁵ whereas the RNN model was based on neuroscience-inspired motion perception models.⁸ Despite inherent architectural differences, we attempted to match the parameters of both networks for

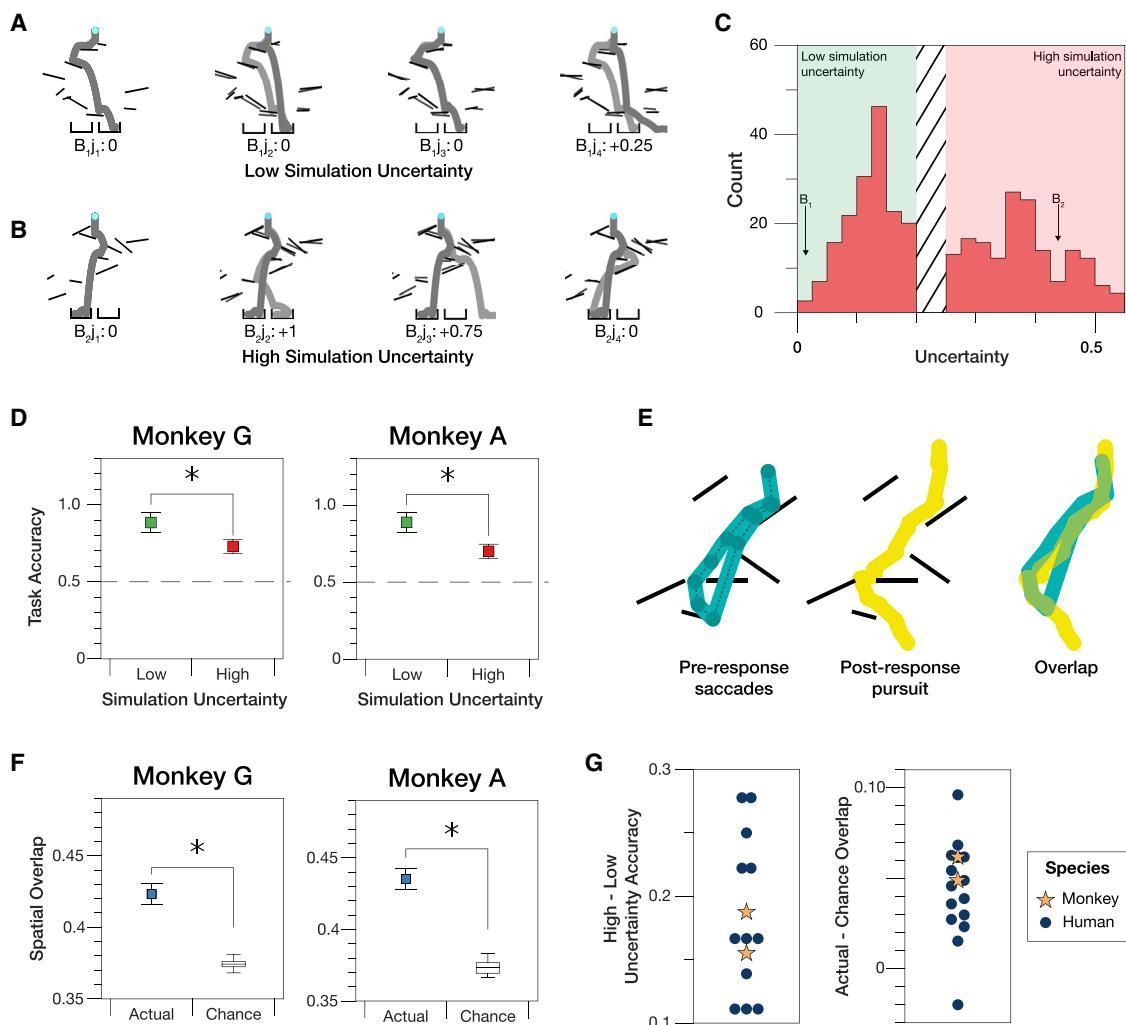


Figure 3. Behavioral methods and evidence for visual simulation in primates

- (A) An example board where slight jittering of plank positions had minimal impact on the ball's outcome (three jittered examples j2,3,4 shown with the original j_1 underlaid), resulting in low simulation uncertainty (see also Figure S3).
- (B) An example where jittering caused significant outcome changes, leading to high simulation uncertainty (three jittered examples j2,3,4 shown with the original j_1 underlaid; see also Figure S3 and Video S2).
- (C) A histogram of simulation uncertainty scores for boards on test days.
- (D) Task accuracy as a function of simulation uncertainty for both monkeys, indicating a possible simulation strategy. Error bars depict standard error of the mean.
- (E) Schematic of eye-movement-overlap analysis between pre- and post-response periods (see also Figure S2).
- (F) Eye-movement spatial overlap relative to chance, consistent with a simulation strategy. Error bars depict standard error of the mean.
- (G) Monkey data compared with human findings, showing similar behavioral and oculomotor trends.

consistency. Figure 4B displays the task accuracy of each network when tested on the same boards as monkey A and monkey G. Like the monkeys, both networks achieved task accuracy greater than chance (CNN: $t_{382} = 15.39, p < 0.001$; RNN: $t_{382} = 25.27, p < 0.001$).

Having established that neural networks could predict the ball's final catcher, we aimed to understand the strategy employed by the two networks by probing the activity of their hidden layers. Figure 4C presents the example activity maps from each network overlaid with the original board. At first glance, the CNN's activity predominantly appears to represent the spatial properties of the planks. Conversely, the RNN seems to focus on the ball's trajectory rather than the planks themselves.

Notably, the RNN naturally emerges with this trajectory representation, despite not explicitly being required or trained to do so. This observation aligns with the behavior one would expect from a system relying on simulation.

To quantitatively confirm these impressions and investigate whether the CNN or RNN represented the ball's trajectory, we trained 16 position decoders to predict equidistant points along the trajectory. We used hidden-layer activity from each network to assess whether these representations contained information about the ball's path. As a control, we repeated the decoding process using the board images, which lack path information (Figure 4D). Comparing decoder outputs with the ground-truth trajectory, we found that several RNN-trained decoders made

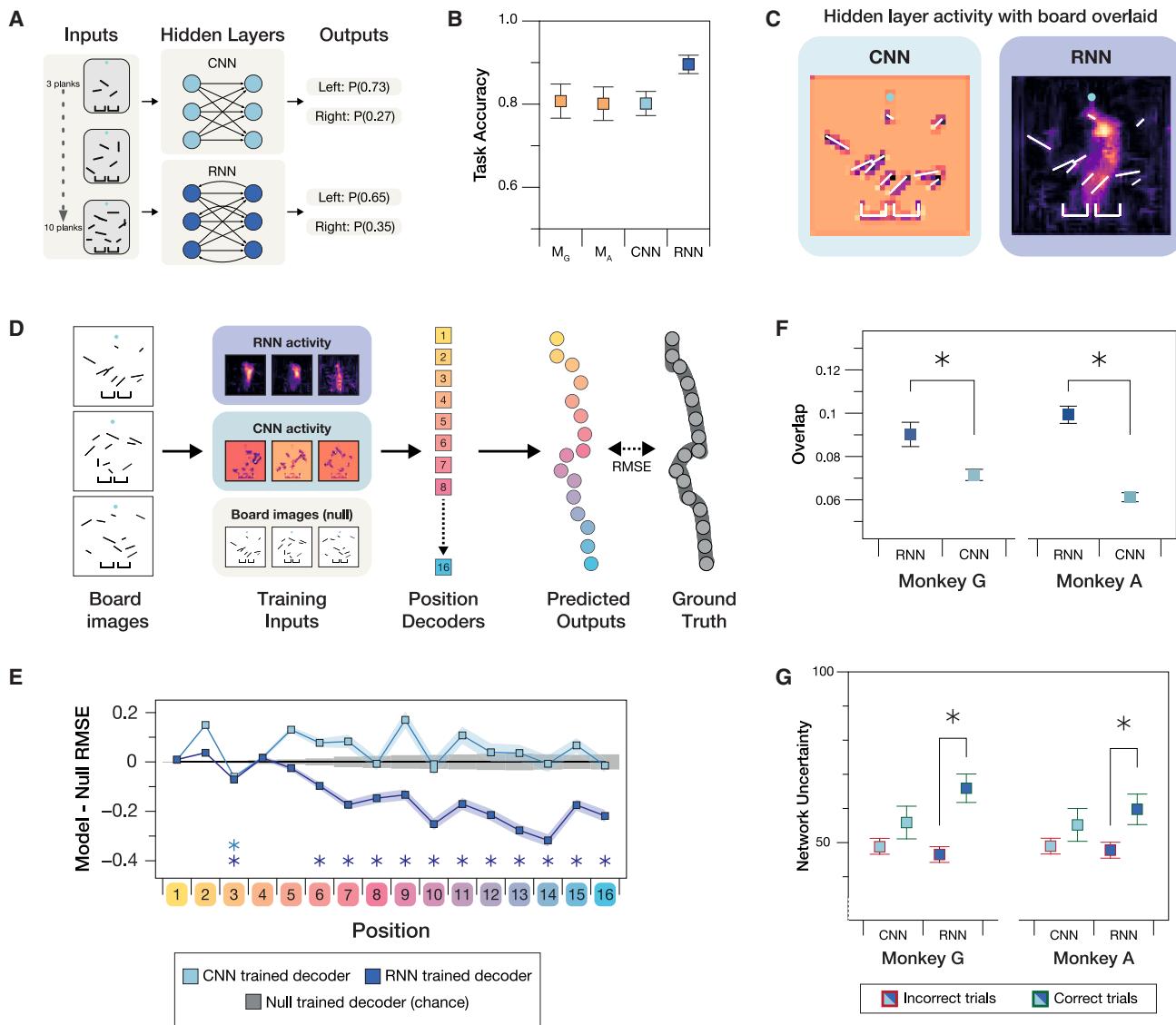


Figure 4. Computational methods and evidence for visual simulation in primates

- (A) Examples of two network types, a convolutional neural network (CNN) and a recurrent neural network (RNN), trained to solve the Planko task.
- (B) Task accuracy of both networks when tested on the same boards from the monkeys' test days. Like the monkeys, both networks achieved above-chance accuracy.
- (C) A heatmap showing the average activity of the hidden units on an example board for both the CNN and the RNN.
- (D) A schematic depicting how we quantified whether the ball's trajectory was represented in the network's hidden-layer activity.
- (E) Average RMSE values for predicted vs. actual positions for CNN- and RNN-trained decoders compared with the null model. Whereas CNN decoders rarely exceeded chance accuracy, RNN decoders consistently predicted the ball's position accurately. See also Table S1 for all statistical tests.
- (F) Spatial overlap between the network's hidden-layer activity and smoothed pre-response eye-movement traces. The RNN had greater overlap than the CNN, suggesting the RNN's simulation-like approach better explains the monkeys' strategy. Error bars depict standard error of the mean.
- (G) Network uncertainty for CNN and RNN as a function of whether monkeys gave correct or incorrect responses. The RNN's uncertainty was significantly higher on boards where monkeys were incorrect, whereas the CNN showed no difference. Error bars depict standard error of the mean.

better predictions than the null-trained decoders, with significant results for 12 of the 16 decoders after Bonferroni correction (Figure 4E; see Table S1). The divergence in prediction ability between RNN and null decoders was most pronounced in the latter half of the trajectory, where variability was highest. By contrast, only one CNN-trained decoder outperformed the null-trained decoders, with most showing similar results (Figure 4E;

see Table S1). These findings confirm that the RNN represents the ball's trajectory, consistent with a simulation-based approach, whereas the CNN adopts a strategy based on the statistical regularities of the planks.

Having established the distinction between each network's hidden-layer activity representation, we compared these activity maps directly to the monkeys' eye movements on the same

board. For each hidden-layer matrix, we filtered for statistically significant values exceeding a critical value at a 0.05 alpha threshold. We then overlaid these activity maps with smoothed traces of the monkeys' pre-response saccades and calculated the spatial overlap (intersection over union) of the two images. This was done board-wise for each monkey and network type. As shown in Figure 4F, the RNN activity maps had significantly greater overlap with the eye-movement traces compared with the CNN maps (monkey G: $t_{190} = -3.37, p < 0.001$; monkey A: $t_{190} = -8.25, p < 0.001$). This result shows that the RNN, which appears to represent the trajectory of the ball (as one might do during a simulation), provides a better account of which areas of the scene the monkeys deemed salient when playing Planko.

Finally, we analyzed whether the boards causing the highest uncertainty for the networks matched those where the monkeys made mistakes. We did this to determine which network approach, simulation-like (RNN) or not (CNN), better aligned with monkey behavior. By calculating each network's board-wise uncertainty (see **STAR Methods**), we found that the RNN displayed significantly higher uncertainty on boards where monkeys responded incorrectly compared with those where they answered correctly (monkey G: $t_{190} = 3.79, p < 0.001$; monkey A: $t_{190} = 2.35, p < 0.05$) (Figure 4G). This result suggests that the monkeys struggled with the same boards for which the answer was unclear to the RNN. Conversely, we found that the CNN exhibited no significant difference in certainty between boards that the monkeys got correct compared with the ones they got incorrect (monkey G: $t_{190} = 1.31, p = 0.18$; monkey A: $t_{190} = 1.21, p = 0.22$), suggesting that the CNN and monkeys struggled with distinct boards. Taken together, these findings indicate that the RNN's simulation-like task approach (as depicted in Figures 3E–4C) better predicts the monkeys' pattern of behavior compared with the CNN's plank-analysis approach, supporting the idea that the monkeys also engaged in simulation.

Neural evidence for visual simulation in monkeys

In previous experiments, we laid the foundation for the idea that monkeys might be capable of simulation. To investigate whether these simulations are visual, we recorded neural responses from monkeys in an MRI scanner while they played Planko. We achieved this by training the monkeys to play Planko in a novel setup—seated in a horizontal chair in the sphinx position. Our hypothesis was that the neural circuits involved in perceiving the falling ball would also be activated during visual simulation. Based on human data, we predicted that motion-sensitive regions would be active during simulation and that this activity would show pattern similarity to when the monkeys perceived the ball falling.

To test these predictions, we defined a motion-sensitive region of interest (ROI) using a motion-localizer task (see **STAR Methods**) in which monkeys fixated on a central spot while white dots either flickered or moved coherently in the background (Figure 5B). Regions showing stronger responses to moving dots, including areas like the MT and the medial superior temporal area (MST) (Figure 6A), were considered motion sensitive. We focused on these motion-sensitive ROIs in subsequent task analyses.

To adapt the task for MRI, we created three Planko variants, administered in blocks (Figure 5C). The first variant, referred to as the "simulation variant," served as the primary experimental condition. In this variant, monkeys were presented with a series of boards and were asked to predict the final catcher for the ball. However, unlike the original Planko task, immediate feedback regarding their choice was not provided. We intentionally removed feedback regarding the ball's trajectory to ensure that monkeys did not perceive any onscreen motion throughout the simulation block. This design choice was crucial to ensure that any activity related to motion perception would not be mistakenly attributed to simulation. Consequently, we introduced a new "perception variant" that served as a positive control. In the perception variant, monkeys were not required to predict the ball's trajectory; instead, as soon as the board appeared on the screen, the ball automatically began to fall. The monkeys were then tasked with retrospectively reporting the catcher in which the ball landed. By examining neural responses from this task variant, we established a template for motion-related activity associated with the ball's trajectory that served as a basis of comparison for the simulation-related activity observed in the previous variant. An additional caveat to note about this variant—to prevent monkeys from "simulating ahead" during this variant, the graphical depiction of the board was made incongruent with the actual ball trajectory. In other words, although we still displayed a normal set of boards as well as a falling ball, the ball would appear to fall through a plank arrangement that was not the one on screen (as seen in Figure 5C).

Finally, we developed a "control variant" to serve as a negative control condition. In this variant, monkeys were presented with static boards (similar to the simulation variant). However, unlike the previous variants, monkeys' responses were no longer dependent on the ball's trajectory. Instead, the orientations of the planks on the screen were manipulated to be predominantly horizontal or predominantly vertical. Monkeys were required to use the orientation property to provide a response (e.g., pressing left for mostly horizontal or pressing right for mostly vertical). Hence, in this variant, monkeys' subjective experience of the task closely resembled that of the simulation variant (i.e., making decisions about a static board during naturalistic free viewing). However, the cognitive processes employed were no longer related to predicting the ball's motion trajectory.

Task variants were blocked, with each run containing 32 trials of a variant (Figure 5D). The monkeys were cued to the run variant by the fixation spot color and generally performed well, achieving 75%+ task accuracy on all runs in the present dataset. The experimental design described here is also consistent with our past human research.⁴ Unfortunately, one of the monkeys developed a fear response to the MRI machine, which meant that only one of the two animals (monkey G) adapted to the altered environment of the scanner. The subsequent results thus all belong to monkey G but are contextualized relative to our previous human findings.

Our results revealed two significant findings. First, within the motion-sensitive ROI, we observed increased activity during the perception variant compared with the control variant (Figure 6B). This finding is expected because the monkeys viewed the falling ball in the perception variant, whereas no motion was present on screen during the control variant. Strikingly,

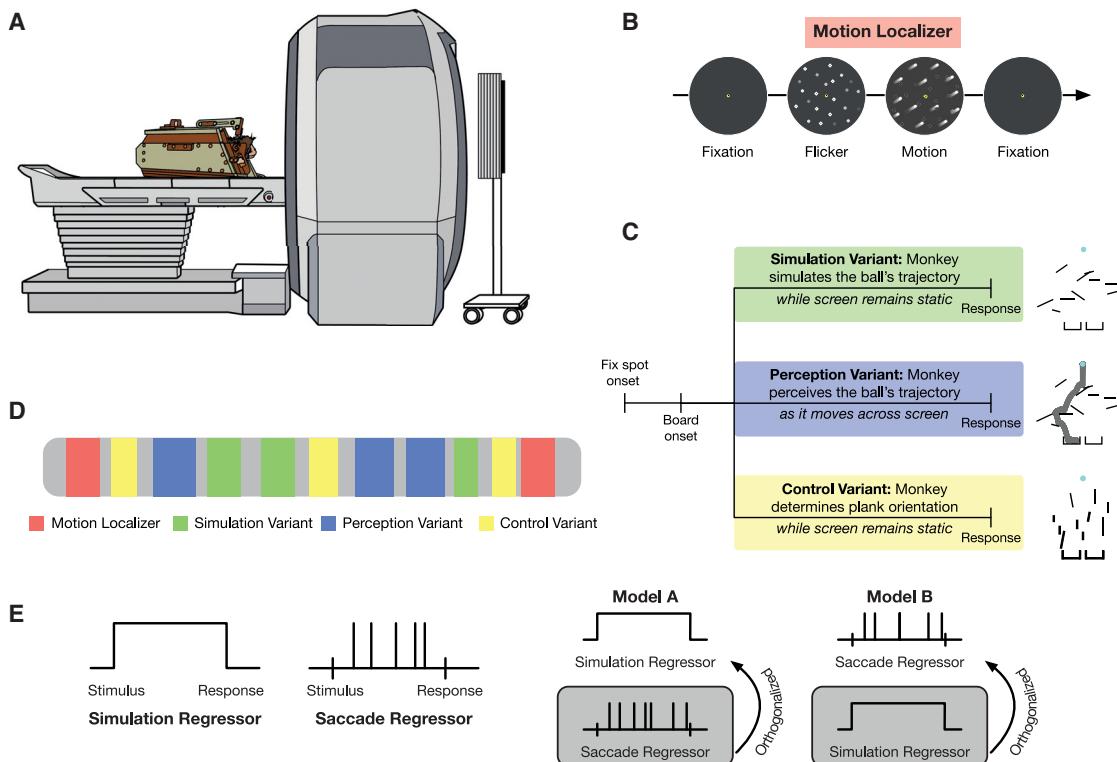


Figure 5. Neural methods for investigating visual simulation

- (A) A diagram of the NHP fMRI setup, where monkeys sat in the “sphinx” position inside the scanner, viewing a screen at the bore’s end and responding via MRI-compatible button boxes.
- (B) A schematic of the motion-localizer task we used to isolate motion-sensitive ROIs.
- (C) A schematic of the three variants of the main Planko task that monkeys were trained to perform inside the scanner.
- (D) An example of one complete scanning session, containing motion-localizer blocks at the beginning and end and several blocks of each task variant randomly interspersed throughout (gray regions indicate interblock intervals).
- (E) Comparison of the “simulation” regressor modeling the pre-response period and a control “saccade” regressor accounting for eye movements. These regressors, once convolved with the hemodynamic response function, were highly collinear. To address this, we ran two control models where the saccade and simulation regressors were orthogonalized, highlighting the unique variance each contributed.

however, we also found increased neural activity in these same motion-sensitive brain regions during the simulation variant relative to the control variant, despite the absence of any onscreen motion in the simulation variant (Figure 6C). The key distinction between the simulation and control variants lies in the fact that the monkeys engaged in a simulation of the ball’s trajectory exclusively during the simulation variant (and only performed orientation discrimination during the control variant). It thus appears that the act of simulating the ball’s path is capable of eliciting activity in motion-sensitive brain areas, as would be expected in the case of visual simulation. Complete activation coordinates for the motion-localizer task and Planko task variants can be found in Tables S2, S3, and S4.

Second, we used representational similarity analyses (RSAs) to compare voxel-level activity patterns between the simulation variant and both the perception and control variants (Figure 6E). Our aim was to see whether the simulation variant’s activity resembled that of the perception variant more than the control variant. Such a pattern resemblance would support the notion that the observed activity during the simulation variant was indeed related to a simulation of the ball’s motion. We found that this was indeed the case, with a higher pattern resemblance

between the simulation and perception variants ($r = 0.45$, $p < 0.001$) compared with the control ($r = 0.2$, $p < 0.001$) (Figure 6G; starred data point). As we only had neural data from one monkey, we assessed the robustness of this finding by conducting several split-half analyses wherein we randomly selected half the runs from each condition and repeated the RSA. We found that across ten data splits, simulation and perception variants were more highly correlated with one another than simulation and control variants, demonstrating the robustness of the finding ($t_9 = -3.6$, $p < 0.01$) (Figure 6F). To further validate this result, we pooled monkey and human data (Figure 6G) and performed several leave-one-out t tests. We observed stable effect sizes r ($d = 1.76$; $SD = 0.14$) across iterations, regardless of the identity of the excluded observation, with all t tests yielding significant differences ($p < 0.001$). These findings suggest that simulation of the ball’s trajectory in monkeys evokes perception-like activity in motion-sensitive brain regions and indicate that, like humans, NHPs may also engage in visual simulation.

It is important to note that the monkey was permitted to freely view the scene as he performed all the task variants, raising concerns that the observed findings might be attributable to

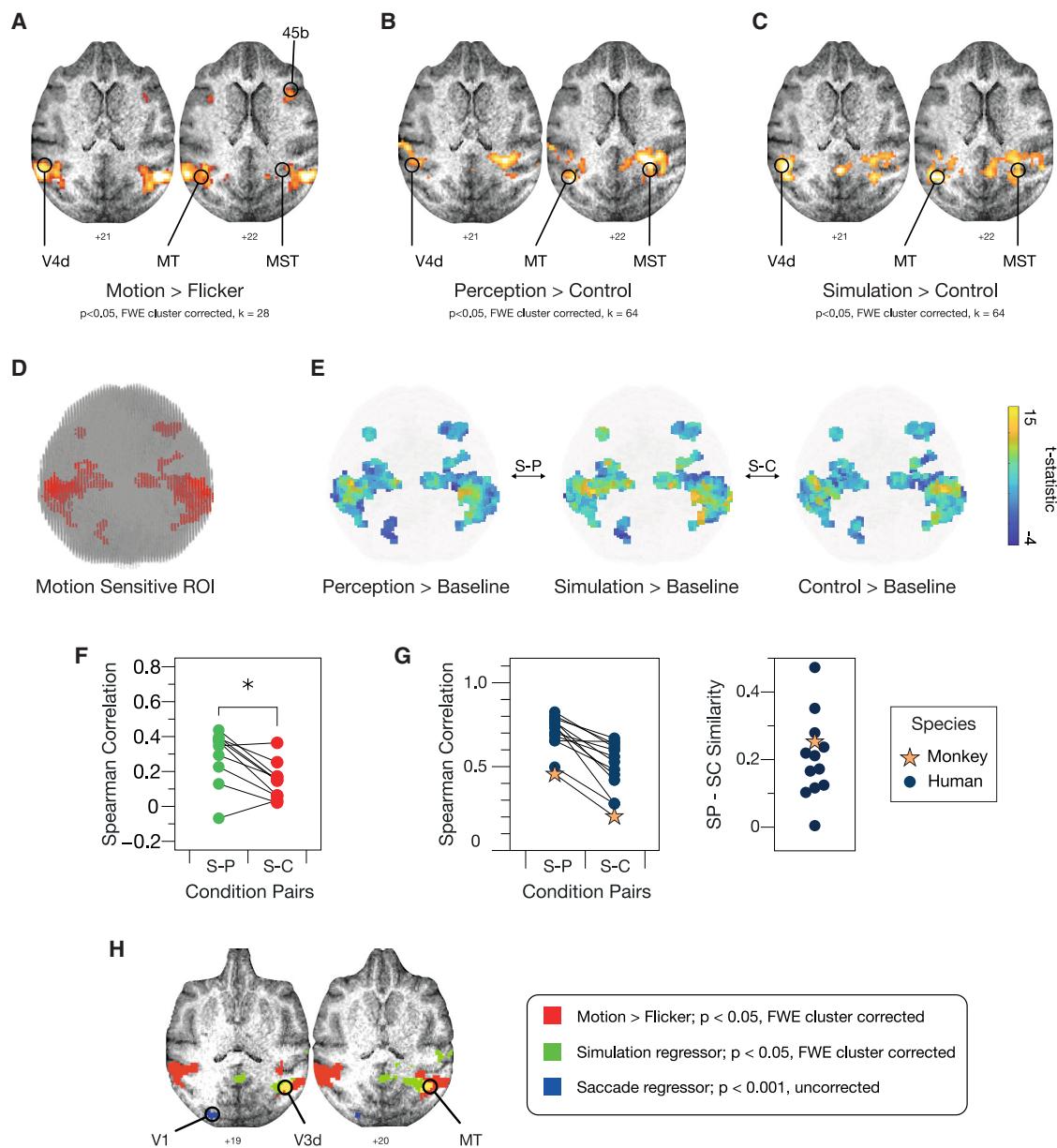


Figure 6. Neural evidence for visual simulation in primates

(A) Results from the motion > flicker contrast in the motion-localizer task showed activity in motion-sensitive brain areas like MT, MST, V4d, and 45b. See also [Table S2](#) for complete activation coordinates.

(B) The perception > control contrast from the Planko task variants also showed activity in MT, MST, and V4d. See also [Table S3](#) for complete activation coordinates.

(C) The simulation > control contrast revealed striking activity in these same motion-sensitive areas. See also [Table S4](#) for complete activation coordinates.

(D) A depiction of the motion-sensitive ROI used for representational similarity analyses. Voxels surviving cluster correction at $p < 0.05$ FWE were selected.

(E) A schematic of key comparisons: simulation vs. perception (S-P) and simulation vs. control (S-C).

(F) Results of 10 split-half analyses showing greater voxel-wise S-P similarities than S-C similarities.

(G) (Left) S-P vs. S-C representational similarities across monkey G's dataset, echoing findings in human participants. (Right) S-P minus S-C similarity in humans and monkey G.

(H) Unique variance from simulation and saccade regressors in two control models showed that saccades explained early visual areas, whereas higher-order motion areas like MT were better modeled by the simulation boxcar. See also [Figure S4](#) and [Table S5](#) for a simulation > perception contrast.

saccade-induced retinal motion. However, past primate fMRI research suggests this is highly unlikely.⁹ Nonetheless, to investigate this possibility in our data, we ran a secondary analysis that included a “saccade regressor” designed to model the monkey’s eye movements. Specifically, we ran two control models: one (model A) with the saccade regressor orthogonalized to the original simulation regressor and another (model B) with the reverse (Figure 5E) to avoid collinearity. We then compared the unique variance explained by each model. We found that the simulation regressor explained variance in clusters overlapping with our motion-sensitive ROI, whereas the saccade regressor only explained variance in early visual areas like V1 (even at a lower significance threshold; Figure 6H). These results align with previous studies and demonstrate that neural activity in the simulation variant in motion-sensitive ROIs is not due to retinal motion.

DISCUSSION

The results of our study offer significant insights into NHP cognition, both behaviorally and neurally. Behaviorally, monkeys demonstrated an impressive understanding of the Planko game, employing strategies beyond basic stimulus-response mappings, suggesting a deeper grasp of the task. Their performance mirrored that of a RNN that represented the ball’s trajectory in the activity of its hidden layers, implying that monkeys were mentally simulating the ball’s movement. Neurally, we observed activation in motion-sensitive brain regions when monkeys mentally simulated the ball’s trajectory, mirroring the neural activity seen when they saw the ball move. This finding aligns with previous human studies, indicating that monkeys, like humans, may engage in visual simulation, underscoring the brain’s ability to predict sensory experiences even in the absence of external stimuli.

Historically, visual simulation—a cognitive process akin to imagination that can be used to predict and plan for the future—has primarily been studied in humans. Nonetheless, the notion that animals might be capable of some form of simulation has gained prominence in recent years. For example, studies on action simulation have suggested that mirror neurons in the motor cortex can internally mimic observed or inferred actions,^{10,11} whereas studies on intuitive physics have demonstrated that when monkeys are asked to intercept a moving virtual ball, their behavior on the task is consistent with a simulation strategy.^{12,13} Mazes have also been used to explore simulation in animals. In one study, spatially tuned neurons in monkeys’ parietal area 7a responded with vectors consistent with the path to a maze’s exit, indicating mental planning.¹⁴ Similarly, rodents navigating a maze have shown hippocampal neural activity that reflects their future path, as if simulating the journey.^{15,16}

Despite these pieces of evidence, definitive conclusions about animal simulation have been difficult to reach due to the complexity and introspective nature of this cognitive process. Moreover, the simplicity of earlier paradigms has often allowed for alternative interpretations of neural activity patterns, such as the recall of past memories rather than active simulation.¹⁷ Our study aimed to address these limitations. For instance, in the Planko task, the configuration of boards on each trial was completely novel, limiting the potential influence of past memory.

The ball’s trajectories were also complex, consisting of several motion vectors, and no information about the path was provided in advance, leaving it entirely up to the monkeys to find the solution. This approach, which is more challenging than previous paradigms, thus strengthens the existing evidence for animal simulation and introduces the possibility that such simulations may have an imagery-like visual component.

The Planko task used in this study relies on an intuitive understanding of physics, and it remains unclear whether monkeys’ ability to learn the task was driven by prior experience or an innate core understanding of the physical properties of the real world. Our choice of a physics-based task was not guided by a desire to emphasize physics; rather, it was to create a system where simulation could be easily transferred between computational and behavioral domains. The physical world we presented was limited to simple rigid-body interactions, but future research could incorporate more complex constraints like joints, gears, or dynamic forces, potentially inducing even more sophisticated internal simulations than those explored here.

If humans and other animals engage in mental simulation, as our data suggest, it raises the question of whether this ability extends to environments not rooted in physical laws. For instance, future tasks could explore simulation in non-physics-based scenarios, including abstract rule-based systems or social interactions. In cognitive science, mental simulation has often been linked to the ability to simulate the thoughts and planned actions of others^{18–20} to predict how others might behave in a particular situation. Although our study focused on predicting the behavior of an inanimate ball in a simplified setting, it is crucial to investigate whether a similar process might underlie monkeys’ ability to anticipate the actions of other animate entities. Understanding this connection could provide key insights into whether and how the brain simulates others’ activities to make sense of the world.²¹

Our use of functional neuroimaging in this and a previous study⁴ offers valuable insights into how people and monkeys visually simulate events. Neuroimaging provides several advantages, including its non-invasive nature and ability to monitor the whole brain. However, it also has limitations. The spatial and temporal constraints of fMRI prevent us from determining the precise sensory details of the neural activation in motion areas. Recordings from individual neurons in these regions could offer a real-time readout of the neural circuits supporting simulation, providing a more detailed understanding of the underlying mechanisms. Nevertheless, our results indicate that mental simulation does indeed activate visual motion-sensitive regions, supporting the idea that monkeys visually simulate imagined events in a way that is consistent with actual perception.

Further, this study focused on asking how “visual” visual simulation is. To that end, we concentrated our neuroimaging analyses primarily on motion-sensitive areas, determined using independent motion localizers. The data provide compelling evidence that simulation of novel motion trajectories does indeed activate these areas and support the naive notion that visual perception of constructed sensory experiences leverages sensory circuitry. Crucially, this did not have to be true. Physics engines capable of solving the Planko task (like the one we used to generate our boards) include no visualization. These engines operate on basic equations for rigid-body interactions in an

entirely computational domain. Our intuitive experience is one of “seeing a simulated world in our mind’s eye,” but this intuition could be a confabulation initiated by running an abstract, non-visual simulation. Our data suggest, though, that we do indeed see things that we imagine during simulation.

Despite our focus here on the visual correlates of simulation, it is important to emphasize that the act of simulation surely relies on coordinated activity across multiple non-visual neural regions as well. How a simulation is initiated, controlled, and monitored remains unclear, but recent work suggests that both prefrontal^{6,22} and parietal cortices^{23,24} play central roles in updating estimates of future states of the world. From our fMRI analyses (Tables S3 and S4), we did find increased activity in temporo-parieto-occipital area (TPO), an association area at the junction of the temporal, parietal, and occipital cortices, in both the perception and simulation variants compared with controls, highlighting the notion that activity in sensory areas represents but a node in the highly complex simulation process.

A significant limitation of the present study was the analysis of neural data from only one animal, despite successfully training two for the Planko task. This issue, although notable, does not substantially detract from the value of our findings, considering the challenges inherent in monkey research and the small sample sizes typically involved. The neural findings observed align closely with past human fMRI research on visual simulation, underscoring the reliability of the results despite the limited dataset. Another limitation is the extensive training required for monkeys to achieve high task accuracy. This raises concerns about the impact of experience on the likelihood of engaging in visual simulation. However, training is an indispensable part of teaching cognitive tasks to monkeys. An interesting future research direction could be to explore whether prolonged practice or long-term exposure influences any neural or behavioral effects related to simulation.

In conclusion, our study using the Planko game demonstrates more than monkeys’ understanding of game mechanics; it reveals their capacity for mental simulation to predict outcomes of events that have not been seen before. This insight is a significant leap in comprehending the relationship between the visual brain and mental experience in NHPs. Moving beyond traditional studies focused on simplified tasks, our findings suggest that animal cognition might encompass complex thought processes akin to human experiences, involving contemplation, simulation, or even “imagination” of potential scenarios. This revelation challenges our current understanding of animal intelligence, indicating that monkeys, and possibly other animals, can weave together past experiences, current observations, and future possibilities. It opens new avenues in cognitive neuroscience, hinting at a rich, imaginative mental landscape in the animal kingdom. As we continue to explore these capabilities, we deepen our understanding of the diverse spectrum of intelligence across species, bridging the gap in our comprehension of cognitive processes in the animal world.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, David L. Sheinberg (david_sheinberg@brown.edu).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- Neural datasets generated in this study will be supplied upon request by the [lead contact](#).
- Any original code that is part of the used analysis will be supplied upon request by the [lead contact](#).
- Any additional information required to re-analyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We would like to acknowledge Michael Worden, Michael Paradiso, Ryan Miller, Kati Conen, and Matthew Maestri for their contributions to this project. We also thank the Brown University Center for Animal Resources and Education for providing animal care as well as the Magnetic Resonance Imaging Research Facility for assisting in implementing scanning procedures. This work was supported by the National Eye Institute, the National Institute of Mental Health, the Office of Integrative Activities, the National Science Foundation, and the Office of Naval Research (grants R01EY014681, R21EY032713, 2T32EY018080-1, 5T32MH115895-02, 1632738, and N00014-24-1-2026, respectively). Additional support was provided by the Carney Institute for Brain Science and its Zimmerman Innovation Fund and the Center for Computation and Visualization (CCV) through the NIH Office of the Director instrumentation grant S10OD025181. We acknowledge the Cloud TPU hardware resources that Google made available via the TensorFlow Research Cloud (TFRC).

AUTHOR CONTRIBUTIONS

A.A. and D.L.S. designed the behavioral experiments. A.A. performed all behavioral experiments and data analyses. A.A., D.L.S., and T.M.D. designed the neuroimaging experiments. A.A. and N.Y.R. performed all neuroimaging experiments and data analyses. A.K.A. and T.S. designed all computational experiments. A.K.A. and A.A. performed all computational experiments and data analyses. A.A. and D.L.S. wrote the manuscript with edits and inputs from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR METHODS

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

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - Subjects and Surgical Procedures
- [METHOD DETAILS](#)
 - Behavioral Experimental Design
 - Motion Localizer and Planko Task Variant Design
 - fMRI Scanning Procedures
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
 - Simulation Uncertainty Analysis
 - Eye Movement Analysis
 - Deep Neural Network Analyses
 - fMRI Data Analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2024.10.026>.

Received: March 23, 2024

Revised: September 3, 2024

Accepted: October 9, 2024

Published: November 15, 2024

REFERENCES

- Springer, A., Parkinson, J., and Prinz, W. (2013). Action simulation: time course and representational mechanisms. *Front. Psychol.* 4, 387. <https://doi.org/10.3389/fpsyg.2013.00387>.
- Battaglia, P.W., Hamrick, J.B., and Tenenbaum, J.B. (2013). Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci. USA* 110, 18327–18332. <https://doi.org/10.1073/pnas.1306572110>.
- Ullman, T.D., Spelke, E., Battaglia, P., and Tenenbaum, J.B. (2017). Mind Games: Game Engines as an Architecture for Intuitive Physics. *Trends Cogn. Sci.* 21, 649–665. <https://doi.org/10.1016/j.tics.2017.05.012>.
- Ahuja, A., Desrochers, T.M., and Sheinberg, D.L. (2022). A role for visual areas in physics simulations. *Cogn. Neuropsychol.* 38, 425–439. <https://doi.org/10.1080/02643294.2022.2034609>.
- Ahuja, A., and Sheinberg, D.L. (2019). Behavioral and oculomotor evidence for visual simulation of object movement. *J. Vision* 19, 13. <https://doi.org/10.1167/19.6.13>.
- Rajalingham, R., Piccato, A., and Jazayeri, M. (2022). Recurrent neural networks with explicit representation of dynamic latent variables can mimic behavioral patterns in a physical inference task. *Nat. Commun.* 13, 5865. <https://doi.org/10.1038/s41467-022-33581-6>.
- Ashok, A.K., Govindarajan, L.N., Linsley, D., Sheinberg, D., and Serre, T. (2022). The emergence of visual simulation in task-optimized recurrent neural networks. SVRHM 2022 Workshop @ NeurIPS.
- Linsley, D., Malik, G., Kim, J., Govindarajan, L.N., Mingolla, E., and Serre, T. (2021). Tracking Without Re-recognition in Humans and Machines. *Advances in Neural Information Processing Systems* 34, 19473–19486.
- Russ, B.E., Kaneko, T., Saleem, K.S., Berman, R.A., and Leopold, D.A. (2016). Distinct fMRI Responses to Self-Induced versus Stimulus Motion during Free Viewing in the Macaque. *J. Neurosci.* 36, 9580–9589. <https://doi.org/10.1523/jneurosci.1152-16.2016>.
- Umiltà, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., and Rizzolatti, G. (2001). I know what you are doing. a neurophysiological study. *Neuron* 31, 155–165. [https://doi.org/10.1016/s0896-6273\(01\)00337-3](https://doi.org/10.1016/s0896-6273(01)00337-3).
- Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. <https://doi.org/10.1093/brain/119.2.593>.
- Rajalingham, R., Sohn, H., and Jazayeri, M. (2022). Dynamic tracking of objects in the macaque dorsomedial frontal cortex. Preprint at Biorxiv.
- Rajalingham, R., Piccato, A., and Jazayeri, M. (2021). The role of mental simulation in primate physical inference abilities. Preprint at Biorxiv.
- Crowe, D.A., Chafee, M.V., Averbeck, B.B., and Georgopoulos, A.P. (2004). Neural Activity in Primate Parietal Area 7a Related to Spatial Analysis of Visual Mazes. *Cereb. Cortex* 14, 23–34. <https://doi.org/10.1093/cercor/bhg088>.
- Basu, R., Gebauer, R., Herfurth, T., Kolb, S., Golipour, Z., Tchumatchenko, T., and Ito, H.T. (2021). The orbitofrontal cortex maps future navigational goals. *Nature* 599, 449–452. <https://doi.org/10.1038/s41586-021-04042-9>.
- Pfeiffer, B.E., and Foster, D.J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* 497, 74–79. <https://doi.org/10.1038/nature12112>.
- Gillespie, A.K., Astudillo Maya, D.A.A., Denovellis, E.L., Liu, D.F., Kastner, D.B., Coulter, M.E., Roumis, D.K., Eden, U.T., and Frank, L.M. (2021). Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *Neuron* 109, 3149–3163.e6. <https://doi.org/10.1016/j.neuron.2021.07.029>.
- Goldman, A.I. (2006). High-Level Simulational Mindreading, Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading, *Philosophy of Mind Series* (Oxford University Press), pp. 147–191. <https://doi.org/10.1093/0195138929.003.0007>.
- Marks, D.F. (1999). Consciousness, mental imagery and action. *Br. J. Psychol.* 90, 567–585. <https://doi.org/10.1348/000712699161639>.
- Dennett, D. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology* (Bradford Books).
- Saxe, R. (2005). Against simulation: the argument from error. *Trends Cogn. Sci.* 9, 174–179. <https://doi.org/10.1016/j.tics.2005.01.012>.
- Fischer, J., and Mahon, B.Z. (2021). What tool representation, intuitive physics, and action have in common: The brain's first-person physics engine. *Cogn. Neuropsychol.* 38, 455–467. <https://doi.org/10.1080/02643294.2022.2106126>.
- Lakshminarasimhan, K.J., Avila, E., Pitkow, X., and Angelaki, D.E. (2023). Dynamical latent state computation in the male macaque posterior parietal cortex. *Nat. Commun.* 14, 1832. <https://doi.org/10.1038/s41467-023-37400-4>.
- Fischer, J., Mikhael, J.G., Tenenbaum, J.B., and Kanwisher, N. (2016). Functional neuroanatomy of intuitive physical inference. *Proc. Natl. Acad. Sci. USA* 113, E5072–E5081. <https://doi.org/10.1073/pnas.1610344113>.
- Leite, F.P., Tsao, D., Vanduffel, W., Fize, D., Sasaki, Y., Wald, L.L., Dale, A.M., Kwong, K.K., Orban, G.A., Rosen, B.R., et al. (2002). Repeated fMRI Using Iron Oxide Contrast Agent in Awake, Behaving Macaques at 3 Tesla. *Neuroimage* 16, 283–294. <https://doi.org/10.1006/nimg.2002.1110>.
- Vanduffel, W., Fize, D., Mandeville, J.B., Nelissen, K., Van Hecke, P., Rosen, B.R., Tootell, R.B.H., and Orban, G.A. (2001). Visual Motion Processing Investigated Using Contrast Agent-Enhanced fMRI in Awake Behaving Monkeys. *Neuron* 32, 565–577. [https://doi.org/10.1016/s0896-6273\(01\)00502-5](https://doi.org/10.1016/s0896-6273(01)00502-5).
- Zhao, F., Wang, P., Hendrich, K., Ugurbil, K., and Kim, S.-G. (2006). Cortical layer-dependent BOLD and CBV responses measured by spin-echo and gradient-echo fMRI: Insights into hemodynamic regulation. *Neuroimage* 30, 1149–1160. <https://doi.org/10.1016/j.neuroimage.2005.11.013>.
- Guo, C., Pleiss, G., Sun, Y., and Weinberger, K.Q. (2017). On Calibration of Modern Neural Networks. In International conference on machine learning.
- Jung, B., Taylor, P.A., Seidlitz, J., Sponheim, C., Perkins, P., Ungerleider, L.G., Glen, D., and Messinger, A. (2021). A comprehensive macaque fMRI pipeline and hierarchical atlas. *Neuroimage* 235, 117997. <https://doi.org/10.1016/j.neuroimage.2021.117997>.
- Grinband, J., Hirsch, J., and Ferrera, V.P. (2006). A Neural Representation of Categorization Uncertainty in the Human Brain. *Neuron* 49, 757–763. <https://doi.org/10.1016/j.neuron.2006.01.032>.
- Grinband, J., Wager, T.D., Lindquist, M., Ferrera, V.P., and Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *Neuroimage* 43, 509–520. <https://doi.org/10.1016/j.neuroimage.2008.07.065>.
- Yarkoni, T., Barch, D.M., Gray, J.R., Conturo, T.E., and Braver, T.S. (2009). BOLD correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fMRI analysis. *PLoS One* 4, e4257. <https://doi.org/10.1371/journal.pone.0004257>.
- Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., and Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* 137, 188–200. <https://doi.org/10.1016/j.neuroimage.2015.12.012>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Macaca mulatta	Mannheimer Foundation Inc	N/A
Software and algorithms		
MATLAB_R2023a	Mathworks	https://www.mathworks.com
AFNI	SSCC at NIMH	https://afni.nimh.nih.gov
SPM12	Wellcome Department of Cognitive Neurology, London, UK	https://www.fil.ion.ucl.ac.uk/spm/software/spm12/
Python	Python Software Foundation	https://www.python.org

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Subjects and Surgical Procedures

Two adult male rhesus macaque monkeys (Macaca mulatta; Monkey A and Monkey G) were included in the study. Both monkeys weighed approximately 15kg and were either solo or pair housed within a larger colony of male monkeys at Brown University. Each monkey was surgically implanted with an MRI-safe Peek headpost to help reduce head motion in the behavioral and MRI experimental setups. Surgeries were performed under isoflurane anesthesia, in accordance with the guidelines published in the National Institutes of Health Guide for the Care and Use of Laboratory Animals. Surgical procedures were approved by the Brown University Institutional Animal Care and Use Committee.

METHOD DETAILS

Behavioral Experimental Design

Both monkeys were trained to perform the Planko task (as described in Figures 2A–2C). Following training, we carried out behavioral test days with each monkey during which they were shown 192 unique boards over the span of 6–8 blocks from 3 sessions. Each trial began with the presentation of a fixation point, following which a Planko board consisting of one ball, ten pseudo-randomly arranged planks, and two catchers was presented on screen. Monkeys had to determine which of the two catchers, left or right, the ball would fall into, were it to be dropped. Monkeys indicated their choice with a button press. The ball was then in fact dropped, revealing the correct answer. Once the ball landed in its catcher, correct responses were rewarded with a few drops of juice. The proportion of boards on which the ball fell into the left or the right catcher was matched (i.e., 0.565 for each). We used an Eyelink-1000 camera (SR Research) to track monkeys' eye movements for the entirety of the session. Eye position was sampled at 1 kHz and stored to disk at 200 Hz.

Motion Localizer and Planko Task Variant Design

Localizer runs started with a 16-second lead-in period with only a yellow fixation point on screen with a black background. Monkeys fixated on the point for the entire 16 seconds. This was followed by randomly ordered 20-second blocks of white dots that either coherently moved in a given direction (i.e., the Motion condition), or flickered on and off (i.e., the Flicker condition). During the Motion and Flicker conditions, the yellow fixation point remained on screen, and monkeys were required to continue fixating (while ignoring the white dots in the background). The white dots were presented in a circular area with a radius of 6 degrees visual angle around the yellow fixation point. White dots were 0.07 degrees visual angle in size and had a density of 69/degrees². During the Motion condition, the white dots moved at 5 degrees/second, randomly changing direction once per second. Monkeys were rewarded for maintaining fixation, which they did for the entirety of each localizer block. Task variant runs (Simulation, Perception, and Control) were broken down into blocks of 32 trials each, starting and ending with a 16-second fixation period. Task variant identity was cued by the color of a fixation spot that was presented at the start of each trial. Monkeys were successfully able to task switch between variant types, even within a single session.

fMRI Scanning Procedures

Monkeys were positioned in an MR-safe chair in the "sphynx" stance, with heads secured using a surgically implanted headpost affixed to the chair's arm. To minimize movement, the chair was padded with Polyethylene foam. Two floor buttons enabled the monkeys to register their task responses. During the task, monkeys wore earplugs to counteract MRI background noise. Prior to scanning, monkeys were administered a contrast agent (MION) intravenously to enhance SNR.^{25–27} Imaging occurred on a Siemens 3T PRISMA MRI system using a custom six-channel coil. Each session began and concluded with a T1-MPRAGE anatomical image repetition time, TR, 2700 ms; echo time, TE, 3.16 ms; flip angle, 9°; 208 sagittal slices; 0.5 0.5 0.5 mm), followed by functional images captured through a fat-saturated gradient-echo echo-planar sequence (TR, 1.8 s; TE, 15ms; flip angle, 80°; 40 interleaved axial slices; 1.1 x 1.1 x 1.1 mm). A 24-inch MRI safe screen displayed the visual stimuli.

QUANTIFICATION AND STATISTICAL ANALYSIS**Simulation Uncertainty Analysis**

To explore simulation uncertainty, we modeled the potential for different ball trajectories on each board by introducing positional jitter to the planks and recalculating the ball's path with a physics engine, as shown in Figures 3A and 3B. Some boards showed significant path deviations with slight plank jitter, while others were unaffected. We used this data to calculate a metric for simulation uncertainty by jittering and recalculating the ball's path 500 times for each board, then measuring how often the jittered configurations resulted in a different outcome. Boards were then classified as low or high uncertainty based on these outcomes, with the results transformed into a 0-100 scale.

Eye Movement Analysis

To assess whether monkeys' eye movements were suggestive of a simulation strategy, we compared their pre-response eye movements (i.e., during hypothesized simulation) to their post-response eye movements (i.e., during perception of the falling ball). It is important to note, however, that eye movements in the pre-response period occurred with a static board presentation, leading to only saccades, while those in the post-response period involved both saccades and smooth pursuit. Due to the distinct nature of saccadic (ballistic) and smooth pursuit (continuous) movements, we did not use traditional oculomotor metrics such as timing and velocity for comparison. Instead, we overlaid the eye movement traces from the pre-response and post-response epochs on top of one another, and then calculated the ratio of the intersection and the union of their areas (Figure 3E).

Notably, this methodology of measuring similarity does result in some incidental spatial overlap even for eye movement traces that are entirely unrelated to one another. We used this form of incidental spatial overlap to quantify a chance intersection level. We did this by randomly shuffling the post-response eye movements across trials and recalculating spatial overlap on mismatched pairs of traces. We implemented this shuffling protocol for each monkey 50 times and averaged the resulting incidental overlap values on each trial for each iteration. Subsequently, we ended up with a distribution of 50 chance overlap values per monkey. We then compared this distribution to the actual, observed degree of overlap between pre-response and post-response eye movements.

Deep Neural Network Analyses

We trained a simple feedforward 2-layer convolutional neural network (the "CNN") and the Index-and-Track (InT) circuit⁸ (the "RNN") each designed to have around 100K parameters. InT incorporates insights from primate neural circuits implicated in object tracking and has been shown to be more performant and correlated to human behavior compared to vanilla RNNs. The RNN consisted of an input layer, the InT circuit layer and finally the readout layer. The input layer had 64 1x1 convolutional filters, the InT circuit had 64 3x3 recurrent kernels mimicking the lateral connections found in the visual cortex. Finally, the readout was a linear layer that transformed the final RNN hidden state to the classification output. The RNN was trained for T = 24 time steps. The CNN was entirely feedforward with a layer of 3x3 convolutional filters followed by a readout layer similar to the RNN.

Both models were trained using the Binary Cross Entropy (BCE) training objective to classify each Planko board into one of either "left" or "right" classes. Model parameters were optimized with Stochastic Gradient Descent implemented via the Adam algorithm (Kingma & Ba, 2014) with an initial learning rate of 3e-4. Planko boards were of size 64 x 64 pixels with 200K boards for training and 5K boards for testing the models. Training was carried out on a NVIDIA TITAN Xp GPU for 100 epochs while measuring validation accuracy after each epoch over a held-out set of 10K boards.

To test if the models had learnt to represent the ball's trajectory, we trained 16 position decoders to predict the position of the ball along the trajectory. For both the CNN and RNN, after training the models to classify the boards, their weights were frozen and the hidden state activities elicited by the 200K training boards were recorded. These activities were fed into a model with three layers of 1x1 convolution and pooling operations and finally a linear layer to obtain the final predicted position. For the control ("null trained decoder"), these same decoders were trained to predict the ball positions directly from the 200K Planko training boards. The decoders were trained to minimize the mean squared error between the predicted ball position and the ground-truth position derived from the physics engine. Like before, the decoders were trained via Stochastic Gradient Descent and were tested on 5K unseen boards.

Finally, we used the networks' uncertainty on each board to ascertain which network's strategy better aligned with monkey behavior. To determine uncertainty, we performed confidence calibration (following training) using temperature scaling.²⁸ This calibrated probability (P(L) and P(R) for "left" and "right", respectively) was used to define the uncertainty for a board as $1 - |P(L) - P(R)|$.

Using this measure on both the CNN and the RNN, uncertainty was calculated for each of the boards on which monkey data was collected (the neural networks were not trained on these boards). Finally, the network uncertainty ascribed to the boards was averaged based on whether the monkeys made an accurate response on said board. That is, we asked if the boards with high network uncertainty scores from a particular neural network were also the ones that the monkeys got incorrect, and vice versa.

fMRI Data Analyses

Data were preprocessed using AFNI (<https://afni.nimh.nih.gov>). Anatomical images were first reoriented, aligned, skull stripped, and warped to the National Institutes of Mental Health Macaque Template version 2.0 (NMT v2) which comes with corresponding Cortical Hierarchy Atlas of the Rhesus Macaque (CHARM) atlas labels.²⁹ Functional images were despiked, slice-time corrected, aligned to the warped anatomical, and smoothed. Volumes with significant motion outliers were censored during the regression analysis. A run was discarded if it was found to require censoring of more than 15% of the total number of volumes. In total, 15 runs per condition passed the criteria for usability, leading to 45 total runs in the dataset.

Task activity on Planko variants was analyzed using a General Linear Model. The expected BOLD response during the pre-trial period was modeled using a boxcar regressor from stimulus onset to participant response. This model was adjusted for varying reaction times, ensuring the accurate representation of the BOLD signal for each trial.³⁰⁻³² The first two trials in each run and trials with outlier reaction times were treated as nuisance regressors. Similarly, nuisance regressors for trials with outlier reaction times, six motion estimates (translation and rotation), and run identity were also included in the model. After these were integrated with the HRF-convolved task regressors, beta and t-statistic values for the task variants were obtained.

After having derived activity estimates for all variants, we conducted a Representational Similarity Analysis (RSA) to compare variants to one another. In this study, we used voxel-wise t-statistics for each variant (contrasted against baseline) within a motion sensitive ROI as the activity estimates due to their demonstrated reliability for RSA.³³ We chose the Spearman correlation as our similarity metric, calculating the degree of similarity between the Simulation and Perception conditions (S-P), as well as the Simulation and Control conditions (S-C). The observed S-P and S-C similarities were then directly compared to one another.