nature communications

Article

https://doi.org/10.1038/s41467-025-61021-8

The capacity limits of moving objects in the imagination

Received: 10 August 2024

Accepted: 6 June 2025

Published online: 01 July 2025

Check for updates

Halely Balaban [1] & Tomer D. Ullman [2,3]

People have capacity limits when tracking objects in direct perception. But how many objects can people track in their imagination? In nine pre-registered experiments (N = 313 total), we examine the capacity limits of mentally simulating the movement of objects in the mind's eye. In a novel Imagined Objects Tracking task, participants continue the motion of animated objects in their mind up to a pre-defined point. When tracking one object in the imagination (Experiment 1a), participants give estimations in line with ground truth. But, when imagining two objects (Experiment 1b), behavior alters substantially: responses are fit best by the predictions of a Serial Model that simulates only one object at a time, as opposed to a Parallel Model that simulates objects in tandem. The serial bottleneck is not due to response/motor limitations (Experiment 2), and is reduced – but not eliminated – by adding extremely strong grouping cues (Experiment 3). Additional studies validate that seriality is found for naturalistic occlusion (Experiment 4) and hyper-simplified physics (Experiment 5), and is not due to factors like noise or lack of motivation (Experiments S1-S3). Altogether, we find that the capacity of moving imagined entities is likely restricted to a single object at a time.

There's only so much we can hold in mind. A well-studied example is the limited ability to track objects in a visual scene. Numerous studies using the Multiple Object Tracking paradigm (MOT[1]) have tested how well people track objects that move about, and found that tracking is limited to a handful of objects[2–5], with ongoing, important debates regarding the exact limitations and their origins[6–9]. These limitations have been examined in great detail in direct perception, but what if the objects are not moving in front of one's eyes, but in the mind's eye? What are the limits of moving objects in the imagination?

People's tracking of objects extends beyond immediate perception, though the exact dynamics of tracking unseen objects or predicting future paths is still debated. In the MOT paradigm, several studies have suggested that people do not extrapolate trajectories to track occluded objects[10,11], at least under most conditions (see ref. 12, for an exception), and instead use heuristics. On the other hand, a main current line of research suggests people use 'mental simulation' to engage in physical prediction or inference, proposing that people continue the trajectories of objects step-by-step in their imagination[13–16]. This approach has accounted for how people reason about the dynamics of objects in a variety of cases[17,18]. While there are ongoing discussions about people's deviation from pure simulation [19–21], here we take as a starting point the idea that people can and do mentally simulate the movement of objects – and use this process to predict, keep track of, and reason about the motion of bodies – but also that this simulation is limited. Given this starting point, our goal was to test whether imagining the future trajectories of objects can be done for more than a single object at a time.

Compared to the large volume of research that examines the capacity limits of processing information available to direct perception, little is known about the limits on tracking imagined objects. While important recent research on mental imagery has started to demonstrate that adding more objects to an imagined static scene increases task difficulty, as reflected in people's subjective reports and precision[22,23], it does not determine the capacity limits of simulating object dynamics in imagination. To examine this, we developed a novel Imagined Objects Tracking task. In this task, people watch animated

[1]Department of Education and Psychology, The Open University of Israel, Raanana, Israel. [2]Department of Psychology, Harvard University, Cambridge, MA, USA. [3]Kempner Institute, Harvard University, Cambridge, MA, USA. e-mail: halelyb@openu.ac.il

scenes in which objects move up to a pause point. People are asked to continue the motion of the objects in their imagination and judge the timing of various outcomes. We focused on timing, as opposed to other dependent measures such as location accuracy (which has been extensively examined and validated in previous work on intuitive physics, but does not determine capacity limits), for two reasons: this avoids imposing serial response requirements, and leads to quantitatively and qualitatively distinct predictions in models of varying capacity.

We compared people's performance in Imagined Objects Tracking to two computational models that implement different hypotheses about the capacity limits of mental simulation (see Fig. 1). According to the Parallel Model, people can mentally advance multiple objects simultaneously. According to the Serial Model, people only advance a single object at a time, unfolding the trajectory of one before going back to unfold the trajectory of another. The Serial Model predicts that every additional object differentially increases the overall imagination-tracking time, delaying people's response for objects that are advanced later mentally. We note that several different sub-types of Serial Models are possible: people might simulate the motion of one object for a number of steps S, then switch to another object, then cycle back again to the first. While a Serial Model that moves each object for a few steps at a time may appear a priori as an appealing solution to how people should mentally simulate objects, we find it completely deviates from the data in all of our studies. Furthermore, while an interleaved model might seem to be a middle ground between a fully serial model and a parallel one, its quantitative predictions do not reflect anything like an averaging of two 'extremes'. Because the interleaved model so clearly does not match our data, and because of its unintuitive predictions, the main text focuses on the serial model that first completely moves one object before turning to the next, but see Supplementary Note 3, and the Discussion, for a complete analysis and consideration of interleaved serial models. We stress that both the Parallel and Serial Models 'keep around' the same number of objects. The capacity limit we studied is with regards to the mental simulation of the dynamics of the objects, and it is not the case that the Serial Model neglects the existence of an object when moving the other forward in time.

There is an important point to emphasize here, which follows a similar debate in the classic tracking literature: finding what looks like a 'parallel' process can be difficult to interpret definitely in favor of parallel processing, since some variant of a serial model can often be constructed to mimic a specific parallel pattern (e.g., rapid switching in a specific way). However, the reverse doesn't hold. Finding a robust serial pattern would provide strong evidence that mental simulation is not done in parallel, and is much harder to interpret under parallel processing. This does not mean that the specific Serial model we used and validated here can capture the full computation in people's minds across all situations. Rather, the simple Serial model tests and validates the unique predictions of a single-item bottleneck, and is the starting point for further Serial models. We return to this point in the Discussion.

In nine pre-registered experiments, we studied the capacity limits of people's ability to mentally simulate the future paths of objects. As a benchmark, we first tested how precisely people track the timing of the imagined trajectory of a single object (Experiment 1a). Next and most important, we examined people's tracking of two objects in the imagination (Experiment 1b), and compared their behavior to the

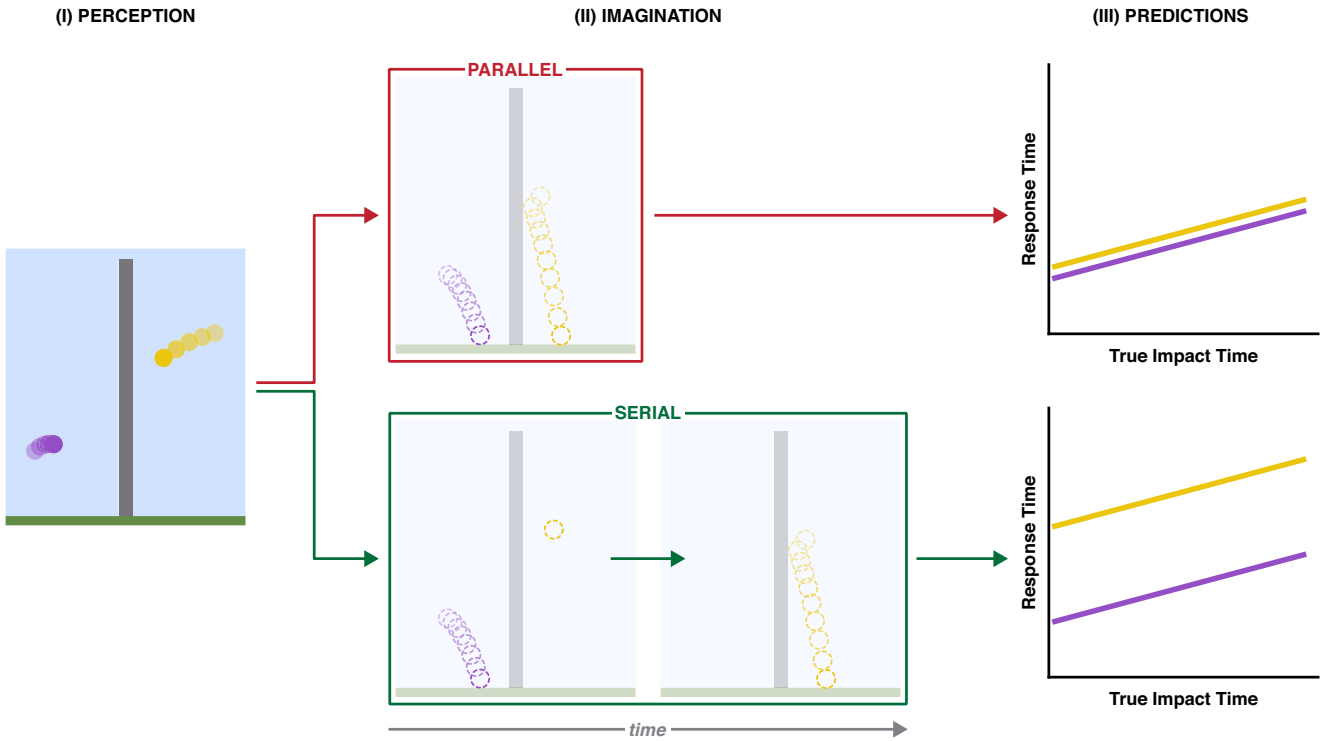

Fig. 1 | Theoretical overview. When watching a scene (I), people can track a handful of objects. But what is the capacity limit of moving objects in the imagination? In the Imagined Objects Tracking task, people watch animations of moving objects that pause mid-motion, and are asked to imagine how the motion continues and estimate the timing of outcomes – here, the moment when each ball hits the ground. Response times (reflecting the subjective impact time) are examined against the actual impact time, which can be manipulated. (II) A priori, moving objects in the mind's eye could happen in Parallel (top), with some number of objects moved forward simultaneously, or Serially (bottom), with only a single object advanced at a time. The Parallel and Serial Models make distinct predictions (III) regarding how people would assess the subjective impact time of objects in a dynamic scene. In the specific example shown in the figure, the Parallel model predicts the subjective impact time of both balls would be roughly the same, while the Serial model predicts a noticeable difference between the first ball moved forward in the imagination (here, the purple ball) and the second (here, the yellow ball).

https://doi.org/10.1038/s41467-025-61021-8

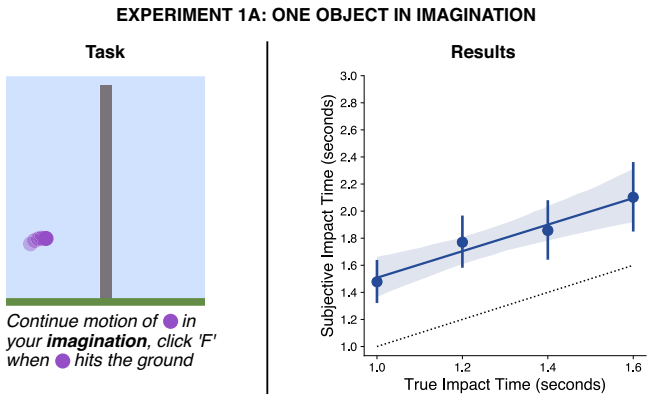

Fig. 2 | Task and Results (n = 36 participants) of Experiment 1a: tracking a single object in the imagination. Circles indicate mean responses for different true impact time, error bars show standard error of the mean (SEM), solid line shows best linear fit, shaded area is 95% confidence interval (CI), and dotted line shows hypothetical perfect performance (where the subjective impact time equals the true impact time), as reference.

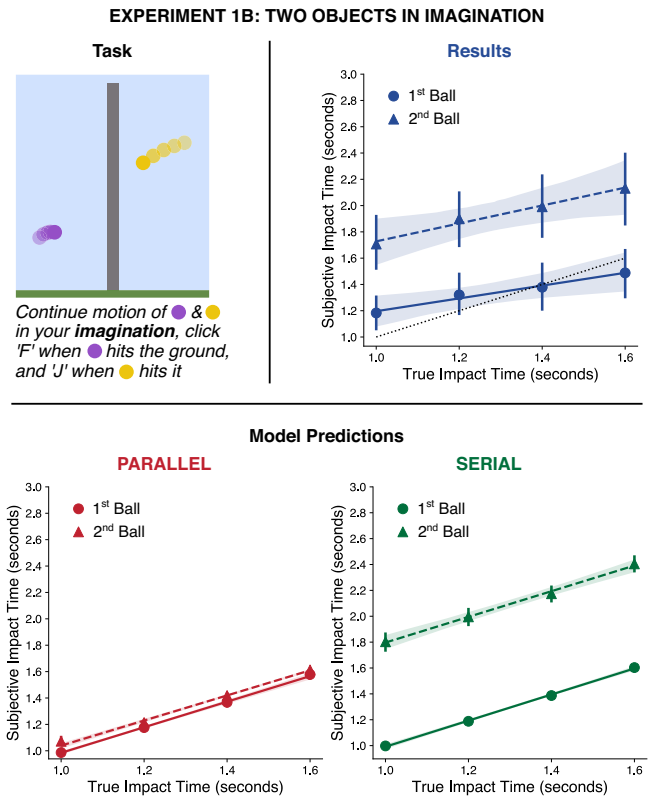

Fig. 3 | Task, results (n = 36 participants), and model predictions of Experiment 1b: tracking two independent objects in the imagination. Circles and triangles indicate mean responses for different true impact time and response order ('1st ball' refers to the ball participants responded to first, and '2nd ball' to the ball they responded to second), error bars show standard error of the mean (SEM), colored lines show best linear fit, shaded area is 95% confidence interval (CI), and dotted line shows hypothetical perfect performance (where the subjective impact time equals the true impact time), as reference. Both models made similar predictions for the first response. For the second response, the Parallel Model predicted a minimal delay, due to perceptual noise. The Serial Model predicted a large delay for the second response, due to ending the simulation of the first object before turning to the second.

predictions of Parallel vs. Serial mental simulation models. We then further examined whether response requirements uniquely contribute to capacity limits (Experiment 2), how scene regularities might help overcoming capacity limits in imagination through grouping (Experiment 3), and how the serial pattern generalizes to naturalistic occlusion (Experiment 4) and simplified physics (Experiment 5). In supplementary experiments (see Supplementary Note 2), we validated that our results are not the effect of noise (Experiment S1), motivation (Experiment S2), or lack of practice (Experiment S3), and additional fine-grained analyses (see Supplementary Notes 1 and 2) supported the same conclusions without aggregating across participants, motion types, or the influence of both items together. Our main finding from these studies is that people's capacity for moving objects in the imagination is extremely limited. Even in the minimal case of continuing the paths of two simple objects, people could only simulate the motion of one object at a time.

Results

Experiment 1a: tracking a single object in the imagination

Participants in Experiment 1a saw animations of a single ball moving according to simulated physics, and pausing mid-motion. They were asked to continue the movement of the ball in their mind's eye and to press a key when the ball (in their imagination) hits the ground (Fig. 2, left). We compared participants' response times—indicating their subjective time estimation—with the actual time it would take the ball to hit the ground, based on the physical simulation. In different animations, the ball moved either like a cannonball or towards the wall, and the true impact time of the ball was manipulated by changing its height and velocity, producing values of 1, 1.2, 1.4, and 1.6 seconds from animation onset (see the Methods section for more details). The goal was to establish whether participants could imagine the future path of a single object in a temporally precise way.

As Fig. 2 (right) shows, responses were linearly modulated by the true impact time (slightly lagging), $F(1.73, 60.49) = 55.01$, $p = 2.4 \times 10^{-13}$, partial $\eta^2 = 0.61$; linear trend: $t(105) = 12.61$, $p = 9.3 \times 10^{-23}$, 95% Confidence Interval (CI) for Mean Difference = [369, 507]. Given that the delay is constant and is not modulated by the true impact time, we take the additive factor to reflect processes unrelated to the imagination component that is our focus, such as motor planning. The linear trend was not an artifact of averaging across participants, and can be seen in the individual data sets of almost all participants (see Supplementary Note 1). The results suggest that people can indeed track the dynamics of a single object in their imagination when they observe scenes like those in our studies. These results serve as the basis for our critical question, which we tackled in the remaining experiments: What happens to people's ability to track objects in the imagination as more objects are introduced.

Experiment 1b: tracking two objects in the imagination

Experiment 1b was identical to Experiment 1a, except that each scene included two objects, and the task was to press a different key when each object hits the ground (Fig. 3, top left). Scenes were created by combining two balls (one moving like a cannonball and one moving towards the wall) from the animations of Experiment 1a, with the true impact time determined independently for each ball, producing a true difference of either 0, 0.2, 0.4, or 0.6 between them. Again, we compared participants' subjective estimation of impact time with the ground-truth impact time, and also broke down the responses by order, meaning the first key press vs. the second one (see also Supplementary Note 1 for an analysis that focuses on the variation in responses instead of the means). Participants overall performed the task well (Fig. 3, top right), with a linear modulation of subjective impact time by true impact time, $F(1.51, 52.86) = 34.99$, $p = 4.8 \times 10^{-9}$, partial $\eta^2 = 0.5$; linear trend: $t(105) = 10.04$, $p = 5 \times 10^{-17}$, 95% CI for Mean Difference = [185, 277] ms. However, the second response happened much later than the first, $F(1, 35) = 103.71$, $p = 5.2 \times 10^{-12}$, partial $\eta^2 = 0.75$. The average delay was 640 ms (95% CI for the intercept of the

first response: [618, 845] ms, second response: [1032, 1405] ms), and the interaction between response order and the true impact time was not significant, $F(2.5, 87.43) = 2.06$, $p = 0.12$. We note that the additive delay in response was smaller than in Experiment 1a, which might reflect a corrective attempt people engage in (i.e., speeding up the simulations to 'catch up' with reality), either explicitly or implicitly. Furthermore, the slope of responses is shallower than in Experiment 1a, suggesting participants are overall less tuned to subtle differences in ground truth physics, likely because of the harder task demands. Because these issues are independent from our main focus on a potential capacity limit in simulation, we set them aside as a target for future research.

We created two simulation models of imagination tracking, using the physics engine that generated the stimuli. The Parallel Model produces two responses that are very close to each other, because both balls are advanced simultaneously (Fig. 3, bottom left). Only a small difference is expected, due to random noise in the simulation, which makes one random ball slightly faster on each run. Conversely, in the Serial Model the first ball has to run all the way through before the second ball can be simulated. Therefore, this model produces a large delay between the first and second responses (corresponding to the first and second ball to be simulated, respectively), in the order of several hundred ms (Fig. 3, bottom right). This is exactly what was found in our participants' data, as reflected by model fits: the Serial Model explained 96% of the variance in average responses, MSE = 0.004, while the Parallel Model explained 20% of the variance, MSE = 0.1.

While the results are rather clear cut in favor of the Serial Model, several concerns present themselves: First, could the serial gap be explained by a very noisy parallel simulation? To test this, in Experiment S1 we independently examined the perceptual noise surrounding object locations with a separate group of participants, and found that it is nowhere near the levels that would be relevant for such a claim.

A second concern is that people may actually be able to carry out a parallel simulation in principle, but simply choose not to in practice, because such a simulation is more effortful than a serial simulation. This concern faces several in-principle difficulties: participants are also presumably motivated by opportunity costs to finish the task quickly, so why not finish it faster through parallel simulation? And why would the total effort of serial simulation over longer periods be less than that of parallel simulation over shorter periods? Beyond such theoretical issues, we empirically tested the motivation concern in Experiment S2, which was similar to Experiment 1b except that we informed participants they would be paid a bonus to the degree to which they were close to the ground truth timing. We found that a motivation manipulation had no effect, and replicated instead the findings of Experiment 1b.

Third, one might wonder whether averaging across the first and second true impact times might obscure important patterns in the results. Specifically, the models can make interesting predictions regarding the influence of the other item's simulation duration (we thank an anonymous Reviewer for this point). The Parallel Model predicts that because the simulations happen in tandem and independently, each response is only affected by that ball's true impact time. The Serial model predicts that the other ball's true impact time should influence the second simulation, because the longer the first simulation takes, the longer the second simulation must wait before it can begin. The fine-grained analysis to examine this requires additional trials, and so we ran a longer version of the task in Experiment S3. First, we replicated the overall serial pattern, suggesting that it is not the result of a lack of task practice. Second, and more importantly, the results followed the serial model's prediction even in terms of the other ball's impact time, which differentially affected the second response but not the first.

A fourth concern may be that averaging across individuals hides important variation, such that some people are able to simulate two or more objects in parallel. However, an individual differences analysis shows this is not the case.

Along somewhat similar lines, a fifth concern is that pooling together across the two movement types we used could obscure important differential patterns, and perhaps even lead to opposing patterns that together average to an illusory serial result (we thank an anonymous Reviewer for this point). Based on this concern, we conducted several post-hoc analyses based on movement type (colliding vs. non-colliding), and found that the serial pattern holds when taking differential trajectories into account.

The full rationale and methods of these additional experiments and analyses are detailed in the Supplementary Information, but to summarize briefly, our conclusion from them and the results of Experiment 1b is that tracking imagined objects via mental simulation is limited to as little as a single object.

We again stress that these results do not mean the simple Serial model we used for predictions captures the full dynamics of people's mental simulation. Rather, the results suggest that mental simulation is not parallel, and works as some kind of serial process. The simple Serial model we considered was sufficient for the findings here, and the full Serial model people use is likely more complex, a point we return to in the Discussion. Still, even if the full Serial model is more complex than the one considered here, such considerations are external to the question of capacity.

Experiment 2: tracking two objects in perception

The results of Experiment 1b suggest an extreme capacity limit in the imagination, such that people only simulated the motion of a single object at a time. However, a major objection is that the bottleneck exists due to a serial response process, instead of in the simulation process. Notably, the requirements of response selection were deliberately minimized: the task involved a constant response mapping, responses were congruent with the side in which each ball appeared, separate hands were used for the two response keys, and participants pressed each key once on each trial. Also, if the bottleneck was such that simulation happened in parallel, but motor-delay caused a constant delay in execution, then we would expect to see a constant additive factor that does not depend on the true impact time of the objects, which contrasts with our findings (for further evidence from individual differences, see Supplementary Note 1).

Still, to more directly test the possibility that the serial bottleneck was created by response execution rather than mental simulation, we conducted Experiment 2. The response requirements of this experiment were identical to Experiment 1b, but the same scenes now played all the way through, meaning participants saw the balls actually hit the ground, without the need to imagine their future paths (see Fig. 4, left). If the serial pattern of Experiment 1b reflects any response-related factor, the results of Experiment 2 should replicate it. But, if the serial pattern is specifically due to the need to simulate the future trajectory of objects, Experiment 2 should be closer to the Parallel Model's predictions.

We found that participants performed well overall, with a linear modulation of subjective impact time by true impact time, $F(1.4, 49.04) = 570.03$, $p = 1.3 \times 10^{-31}$, partial $\eta^2 = 0.94$; linear trend: $t(105) = 41.3$, $p = 9.5 \times 10^{-67}$, 95% CI for Mean Difference = [324, 358] ms. As can be seen in Fig. 4 (right), instead of replicating the serial pattern of Experiment 1b, the results of Experiment 2 revealed a much smaller response delay. The second responses were slower, $F(1, 35) = 60.15$, $p = 4.1 \times 10^{-9}$, partial $\eta^2 = 0.63$, in a way that now interacted with true impact time, $F(2.36, 82.62) = 31.74$, $p = 4.9 \times 10^{-12}$, partial $\eta^2 = 0.48$, due to a larger effect for smaller impact times. Critically, the effect of response order was smaller than in Experiment 1b, $F(1, 70) = 74.65$, $p = 1.2 \times 10^{-12}$, partial $\eta^2 = 0.52$. As can be seen also in the Parallel

https://doi.org/10.1038/s41467-025-61021-8

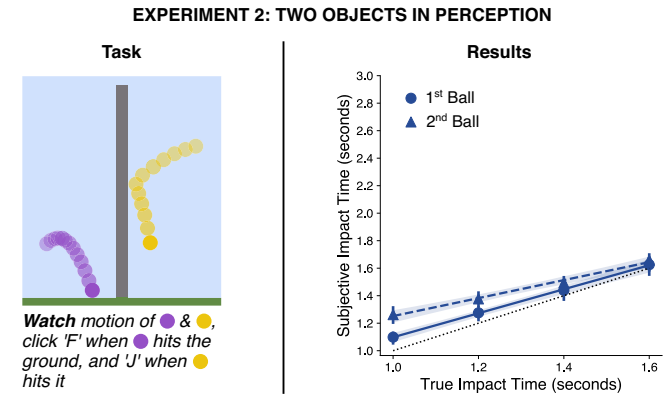

Fig. 4 | Task and results ($n = 36$ participants) of Experiment 2, tracking two objects in perception. Circles and triangles indicate mean responses for different true impact time and response order ('1st ball' refers to the ball participants responded to first, and '2nd ball' to the ball they responded to second), error bars show standard error of the mean (SEM), colored lines show best linear fit, shaded area is 95% confidence interval (CI), and dotted line shows hypothetical perfect performance (where the subjective impact time equals the true impact time), as reference.

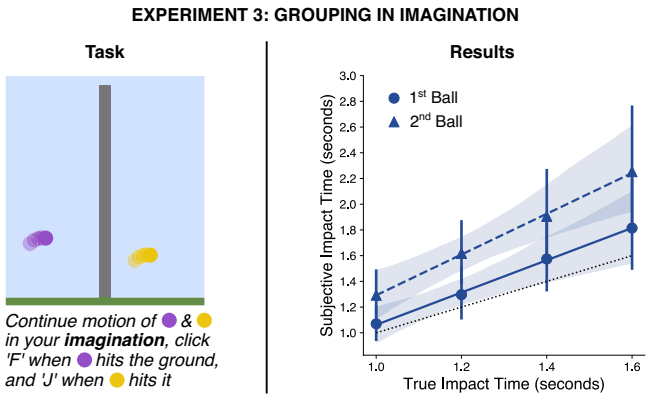

Fig. 5 | Task and results ($n = 36$ participants) of Experiment 3, tracking two objects in the imagination with strong grouping cues. Circles and triangles indicate mean responses for different true impact time and response order ('1st ball' refers to the ball participants responded to first, and '2nd ball' to the ball they responded to second), error bars show standard error of the mean (SEM), colored lines show best linear fit, shaded area is 95% confidence interval (CI), and dotted line shows hypothetical perfect performance (where the subjective impact time equals the true impact time), as reference.

Model's predictions, some effect of response order is always expected (by definition, the second response is slower than the first), but as Experiment 2 empirically shows, the effect is largely reduced, averaging at 88 ms (95% CI for the intercept of the first response: [181, 239] ms, second response: [514, 678] ms). Model fits confirmed that the Parallel model was preferred for tracking in perception: the Parallel Model explained 97% of the variance in average responses, MSE = 0.001, while the Serial Model explained 47% of the variance, MSE = 0.02.

We stress that we do not take the results of Experiment 2 to definitively reflect either parallel or serial perceptual tracking. While the results are more aligned with the Parallel Model, that model refers to mental simulation, and it is possible that in perception people are either carrying out the task in parallel, or through very rapid serial switching. Whether it is one or the other does not matter to our central point here: the results of Experiment 2 differed drastically from Experiment 1b, and show that the serial pattern of Experiment 1b are not due to a bottleneck in response requirements (which were identical in Experiment 2). Instead, the results likely reflect a specific serial constraint on simulating the paths of imagined objects.

Experiment 3: tracking two objects in the imagination with grouping

The finding that people mentally simulate a single object at a time is striking when considering the simplicity of the current task compared to real-world tasks, which regularly involve many objects that can move in complex paths. Mental simulation likely evolved to employ different hacks[16] that might overcome the serial bottleneck found here. One important strategy could leverage regularities in the environment, such as Gestalt cues, which improve performance in perceptual tracking[24]. It seems reasonable to expect that if the motion paths of different objects are similar enough, the objects will be grouped in imagination, allowing their physics to be advanced in parallel. We tested this idea in Experiment 3, which used the same imagination task as in Experiment 1b, but with three important modifications (see Fig. 5, left): only hyperbole motion was used, the two balls always moved in the same direction (either to the left or right, instead of toward each other), and velocity was held constant. This meant that the visible motion sequence was identical for all items, to encourage participants to group the two balls in each scene. Because the true impact time was determined solely by a ball's initial height, the setup also created a greater opportunity for using heuristics instead of imagining the exact trajectory, which could be another way of overcoming the single-item capacity limit.

Participants performed the task reasonably well, and the subjective impact times were linearly modulated by true impact time, $F(1.05, 36.8) = 36.84$, $p = 3.5 \times 10^{-7}$, partial $\eta^2 = 0.51$; linear trend: $t(105) = 10.51$, $p = 4.3 \times 10^{-18}$, 95% CI for Mean Difference = [515, 755] ms. As can be seen in Fig. 5 (right), the second responses were still much slower than the first, $F(1, 35) = 62.65$, $p = 2.6 \times 10^{-9}$, partial $\eta^2 = 0.64$, in a way that interacted with true impact time, $F(2.16, 75.84) = 4.1$, $p = 0.02$, partial $\eta^2 = 0.1$, this time because of a larger difference for larger impact times. The average difference between the first and second responses was 328 ms (95% CI for the intercept of the first response: [−461, 47] ms, second response: [−670, 28] ms), which was smaller than in Experiment 1b, $F(1, 70) = 17.13$, $p = 9.6 \times 10^{-5}$, partial $\eta^2 = 0.2$, but larger than in Experiment 2, $F(1, 70) = 31.17$, $p = 4.2 \times 10^{-7}$, partial $\eta^2 = 0.31$. Accordingly, we found an intermediate pattern based on the fit between participants' data and our computational models: the Serial Model explained 69% of the variance in responses, MSE = 0.04, and the Parallel Model explained 86% of the variance, MSE = 0.02. The results suggest that grouping could relax the single-item bottleneck of imagination tracking somewhat, but not eliminate seriality completely, even with identical motion sequences and an opportunity to use heuristics.

Experiment 4: tracking two objects in the imagination with natural occlusion

So far, our studies suggest that simulating the movement of objects in the imagination is limited to a single object, and multiple objects are simulated serially. Our next two experiments examine the generalization of this conclusion beyond the specifics of the stimuli and tasks used in the previous studies.

One important issue to test was whether the capacity limits we observed were related in some way to a disruption of tracking, due to the objects freezing in mid-air instead of disappearing in a more ecological way (see ref. 25, we thank an anonymous Reviewer for pointing this out. So, in Experiment 4 we repeated the task and stimuli of Experiment 1b, except that the stimuli included an occluder (see Fig. 6, left). The dimensions and placement of the occluder were chosen so

https://doi.org/10.1038/s41467-025-61021-8

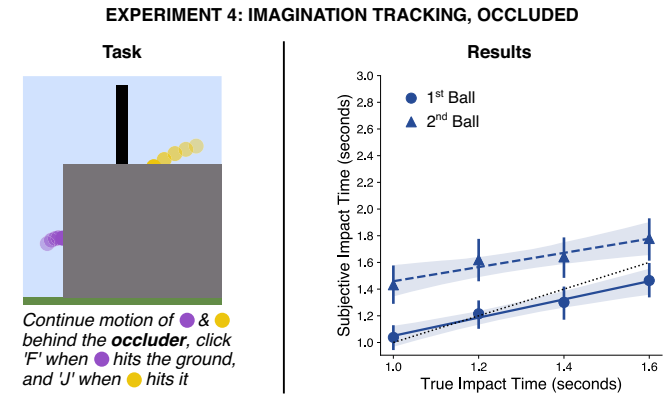

Fig. 6 | Task and results ($n = 36$ participants) of Experiment 4, tracking two objects in the imagination with natural occlusion. Circles and triangles indicate mean responses for different true impact time and response order ('1st ball' refers to the ball participants responded to first, and '2nd ball' to the ball they responded to second), error bars show standard error of the mean (SEM), colored lines show best linear fit, shaded area is 95% confidence interval (CI), and dotted line shows hypothetical perfect performance (where the subjective impact time equals the true impact time), as reference.

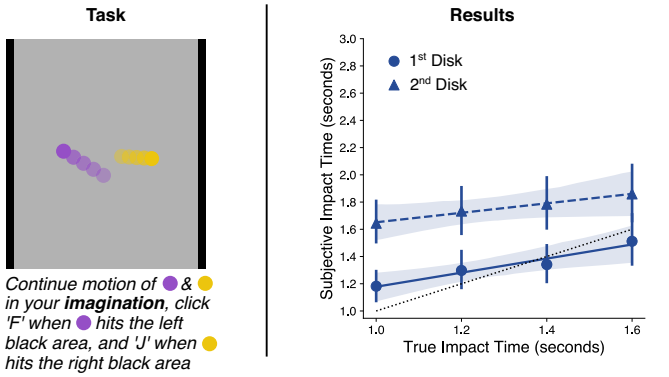

Fig. 7 | Task and results ($n = 36$ participants) of Experiment 5, tracking two objects in the imagination with minimally physical dynamics. Circles and triangles indicate mean responses for different true impact time and response order ('1st disk' refers to the disk participants responded to first, and '2nd disk' to the disk they responded to second), error bars show standard error of the mean (SEM), colored lines show best linear fit, shaded area is 95% confidence interval (CI), and dotted line shows hypothetical perfect performance (where the subjective impact time equals the true impact time), as reference.

that the objects disappeared behind it after 500 ms of movement, in keeping with the previous studies.

As shown in Fig. 6 (right), the results of Experiment 4 replicate the results of Experiment 1b, demonstrating that mentally simulating items that disappear in a natural way behind an occluder produces a serial pattern. Participants' responses were linearly modulated by the true impact time, $F(2.02, 70.85) = 158.91$, $p = 4.8 \times 10^{-27}$, partial $\eta^2 = 0.82$; linear trend: $t(105) = 21.43$, $p = 3.9 \times 10^{-40}$, 95% CI for Mean Difference = [246, 296] ms. As in the experiments that involved items freezing, the second response happened much later than the first, $F(1, 35) = 141.28$, $p = 7.6 \times 10^{-14}$, partial $\eta^2 = 0.8$. Comparing the results to Experiment 1b, we found a significant interaction of Experiment with Response Order, $F(1, 70) = 15.65$, $p = 1.8 \times 10^{-4}$, partial $\eta^2 = 0.18$, driven by a larger effect in Experiment 1b. In Experiment 4, the average delay was 364 ms (95% CI for the intercept of the first response: [226, 509] ms, second response: [759, 1097] ms), and the interaction between response order and the true impact time was significant, $F(2.36, 82.64) = 4.78$, $p = 0.007$. Examining model fits showed that the Serial model explained 96% of the variance in responses, MSE = 0.002, while the Parallel model explained 46% of the variance, MSE = 0.03.

The results of Experiment 4 suggest the serial capacity limit is not due to the freezing of the objects, and is observed also for items that move more naturally behind occluders. We note that occlusion did mitigate the serial effect somewhat, and this might reflect any of a number of factors, like the greater predictability of when objects will disappear behind the occluder, as opposed to when they will freeze, or the greater familiarity the ecological motion of moving into occlusion compared to freezing in mid-air. Overall, given how common occlusion is in the real world, the replication of the serial pattern in Experiment 4 suggests that the single object bottleneck arises under naturalistic simulation conditions as well, corroborating the importance of the present findings.

Experiment 5: tracking two objects in the imagination with minimally-physical scenarios

Another issue to test was whether there is something uniquely complicated about a situation involving realistic physics of objects falling under gravity, as opposed to the more simplified stimuli often used in MOT. We note that a priori, this seems unlikely. The simplified 2D scenes we used can hardly be considered complex, and if anything, we might expect people to be better adjusted to the more ecological task of objects colliding and falling under gravity than the not-frequently-encountered task of objects moving in free-form, as in many MOT studies.

Still, it is worth bringing our stimuli more in line with many MOT studies, and examining whether the serial pattern holds for even more simplified simulations. Participants in Experiment 5 performed an imagination tracking task similar to the one used in the previous studies, but the scenes were altered so that the stimuli were similar to many MOT tasks (see Fig. 7, left): gravity was turned off, collisions were eliminated, the background was gray with two black rectangles on either side, and two colored disks moved in straight lines and with a constant speed from roughly the center of the screen to the right and left sides of the screen.

As shown in Fig. 7 (right), the results of Experiment 5 replicate the results of Experiment 1b. Participants' response were linearly modulated by the true impact time, $F(1.41, 49.40) = 26.44$, $p = 3.5 \times 10^{-7}$, partial $\eta^2 = 0.43$; linear trend: $t(105) = 8.79$, $p = 3 \times 10^{-14}$, 95% CI for Mean Difference = [149, 237] ms. However, as in the more physical experiments, the second response happened much later than the first, $F(1, 35) = 132.56$, $p = 1.8 \times 10^{-13}$, partial $\eta^2 = 0.79$. Comparing the results to Experiment 1b, we found a significant interaction of Experiment with Response Order, $F(1, 70) = 8.92$, $p = 0.004$, partial $\eta^2 = 0.11$, driven by a larger effect in Experiment 1b. In Experiment 5, the average delay was 423 ms (95% CI for the intercept of the first response: [442, 880] ms, second response: [1075, 1532] ms), and the interaction between response order and the true impact time was not significant, $F(2.37, 83.15) = 2.07$, $p = 0.12$). In terms of fits, the Serial model explained 98% of the variance in responses, MSE = 0.001, while the Parallel model explained 27% of the variance, MSE = 0.04.

These findings demonstrate that our results are not due to the specific physical requirements imposed by the main task. This suggests that tracking two objects in the imagination using minimally-physical stimuli that are similar to conventional perceptual tracking tasks is done serially. Mentally simulating the movement of items appears to happen on a single-item basis for both more complex and minimal stimuli.

Discussion

Research spanning decades has demonstrated that people have signature capacity limits when tracking visible objects. Here, we

examined capacity limits when objects were moving in the imagination. We found that the mind's eye can only track a single object at a time. More specifically, we found that people could reasonably unfold the trajectory of a single object in the imagination (Experiment 1a), but that the addition of just one independent object substantially altered their responses (Experiment 1b), in line with the predictions of serial mental simulation. This Serial Model suggests people first had to mentally advance one object up to some point, before going back and advancing the second object. We did not observe the capacity bottleneck when people tracked two objects in perception instead of imagination (Experiment 2), further cementing the notion that the capacity limit is in mental simulation, not the motor response or other limits further downstream. Additional experiments, models, and analyses (see the Supplementary Information) showed that the serial gap is not the result of noisy parallel simulation, lack of motivation, lack of practice, or averaging across different motion dynamics, and also that the limitations hold at the individual participant level. Notably, the stable serial pattern emerged despite of the well-known difficulty of observing serial costs in performance (the difficulty of teasing apart serial and parallel patterns applies when a seemingly parallel pattern could be interpreted as very rapid serial switching, but that is not the case here). When we added strong grouping cues to the trajectory of the objects, we found that the difference between the first and second response shrank, but was not fully eliminated (Experiment 3). Finally, the single-object bottleneck generalized to other stimuli and tasks, like when unnatural freezing was replaced by ecological occlusion (Experiment 4) and when items moved in straight lines without any aspect of complex physics (Experiment 5). Taken together, our results suggest that mentally simulating the movement of objects is a serial process.

The finding that people are able to track only a single object at a time in their imagination is surprising, seeing as people can usually track a handful of items in direct perception (though the exact number is affected by various factors, such as object speed or spacing[1,2,5,8]). If seeing things in the mind's eye is supposed to be akin to seeing with one's real eyes[26], our findings suggest it isn't so. However, while researchers do use the term 'track' to include the following of hidden objects behind occluders in perceptual tasks, it may be that this is an over-loaded term. 'Tracking' in the imagination (or through occlusion) may be quite different than direct perceptual tracking, as it is the mind itself that is moving the objects, rather than keeping on top of objects that are being moved by external forces. Such a distinction aligns well with two lines of work in attention and working memory. First, people's ability to extrapolate motion in perceptual tracking was recently suggested to have a capacity limit of only one object[6], perhaps due to challenges of physical simulation[27]. Second, updating active representations was argued to depend on sequentially loading objects, one at a time, into the 'focus of attention'[28]. So, it may be that calculating an object's future motion (whether in direct perception as in MOT, or in imagination as in Imagined Objects Tracking) requires constantly updating the object's representation in working memory, and that relies on a serial process (for additional connections between physical simulation and working memory, see ref. 29. This is further strengthened by our finding that the serial pattern can be observed not only for items that freeze mid-motion, but for items that undergo natural occlusion (Experiment 4). The capacity limits we found in the imagination should then be taken to refer to the simulation part that moves objects forward, rather than to a later stage that re-processes the imagined scene. Also, we did not control for eye movements, and it is possible that people followed imagined trajectories with their eyes, which contributed to the single-object bottleneck, although this only raises the question of why people could not shift their eyes to track both imagined objects (as they do in perception). This is not a limitation of the studies, but rather is in line with how people may carry out physical predictions[30–32], and is an interesting topic for future research.

Independent of capacity limits in perception, another reason why our findings are surprising is that they contrast with subjective intuitions about internal scenes. Many people report being able to imagine vividly dynamic mental scenes, and a single-object capacity in simulation doesn't align with that. Why then does it subjectively seem like we can imagine vividly moving dynamic scenes? This is similar to the apparent conflict between our intuition and other capacity limits – for example, in our everyday life, we do not feel like we only have access to a tiny subset of all of the perceptual input, yet decades of working memory research have shown that this is indeed the case, and under naturalistic conditions we simply rely on other mechanisms for compensation, such as long-term memory or scene scanning with saccades (for a review of similar ideas, see ref. 33). Aside from general arguments regarding the unfaithful nature of introspection[34], our third study showed that grouping does ameliorate the serial effect (though it doesn't negate it). Our main effect relied on intentionally creating scenes in which mental objects move independently of one another, but this may not be a typical case. It is likely that many dynamic mental scenes (perhaps also those previously used in intuitive physics research[15,35]) rely on strong grouping and hierarchical organization[36], such that the serial process need only update a hyper-parameter that controls the motion of several objects at the same time. One such example would be mentally simulating the distance between items, instead of the location of each item separately, an interesting idea that can be the target of future work.

Our studies focused on non-interacting objects to keep the findings clear and simple, but objects in the mind can interact. A simple case study of minimal interaction is two objects moving along a plane at various speeds and angles, possibly about to collide. In such a case, it seems unlikely that people use a serial updating process that fully moves the first object, then the second, as no collisions would occur. In such a case, perhaps people move forward one object for a limited number of steps S, then switch to another object, and cycle back again. As detailed in Supplementary Note 3, such 'interleaved' serial simulation models did not explain the data in the present studies, but they may be relevant for interaction/collision situations, possibly with a dynamically set S. We plan to pursue such cases and models in future studies.

The complications introduced by interacting objects relate to another important point: even though the plain Serial Model captures almost all the variation in people's responses in our experiments, mental simulation is almost certainly more complex than this model. We think of our results as strongly speaking against a parallel mental simulation across a range of situations, but not as specifying a complete model of all aspects of people's precise serial simulation. For example, a full model of physical mental simulation must account for how people resolve collisions (which is known to be challenging and noisy[37]), or how people employ heuristics regarding which item to simulate first. We briefly begin touching upon these issues in the Supplementary Information (see Supplementary Note 1), but the main point is that these interesting questions are external to our core investigation here. The present study revolves around the number of items that can be simulated at once, which we found to be one, but leaves open questions regarding the precise way this one item is simulated.

Another intriguing direction for future research concerns the information that people do manage to simulate. Specifically, it is unclear whether people imagine the objects along with all of their features, or are closer to a computerized physics engine that handles only trajectories[38]. The current results cannot offer an answer to this question, and it is independent from the issue of the capacity limit of the simulation process. Yet, past findings from MOT do point to a differential status of spatiotemporal information and surface features at least in perceptual tracking. An extreme manifestation of this is that while featural information can definitely aid tracking by allowing for

more efficient deployment of attentional resources[39], when the tracked objects change their features, people might entirely miss this[40]. On the other hand, MOT findings suggest that people manage to rely on featural information for grouping[24], and so it would be interesting to test whether the single object capacity limit in mental simulation might be relaxed not only by physically-relevant information (as the identical motion paths used in Experiment 3), but also by presenting the objects in the same color.

The capacity limits we found hold independently of the specific cognitive computations one assumes people use to advance objects in the mind's eye. That said, we do adhere to a mental simulation approach to intuitive physics[14], and our computational models did assume that people track objects in the imagination by mentally advancing them step-by-step. This view contrasts with research that argues that humans do not rely on mental simulation for intuitive physical judgments, and which often points to people's systematic mistakes and deviations from ground-truth physics as evidence against simulation. While these two views are often portrayed in opposition, we see the current work as another brick in the bridge between the rich literature on errors in physical judgments[20,41] and the mental game engine framework. It is part of a general approach that uses game engines as inspiration for an overall mental simulation account, but also draws on the shortcuts and workaround used in such engines to save on time, memory, and overall computation[16]. Such an approach has found evidence for people's use of systematic approximations in the representations of bodies themselves[42], as well as people's use of 'partial simulation', in which they do not mentally simulate parts of the scene that are deemed irrelevant[21]. Our present work shows another central way in which mental physics parts ways with real physics, while still being overall consistent with a mental simulation account.

We set out to examine the capacity limits of the imagination. We found that even in a simple situation the answer to 'how many objects can the mind's eye keep track of at once?' is 'approximately one'. This might feel like discovering you've been tricked. Like realizing that who you took to be a fantastic juggler is really only tossing and bouncing a single ball. Still, knowing the trick makes you appreciate the act in a different way: It's poor juggling, but it's a great trick.

Methods

Materials, data, code, and pre-registration protocols for all experiments are available in the following Open Science Framework repositories: https://osf.io/wzt98/ and https://osf.io/kcmbs/; the protocols were pre-registered on April 5th 2024 (Exp. 1a, 1b, 2, 3, and S1), July 7th 2024 (Exp. 5 and S2), November 13th 2024 (Exp. 4), and February 27th 2025 (Exp. S3). Studies ran tasks coded in jsPsych (version 7.2.3). Data was collected using JATOS (version 3.8.6) hosted on MindProbe, and analyzed using custom Python (version 3.9.6) codes that are available online as detailed below.

Participants

Research was approved by the Harvard University Ethics Committee (protocol IRB19-1861), including all studies reported in both the main text and Supplementary Information. All participants, including in the Supplementary studies, provided informed consent. Participants were recruited online[43] via Prolific (https://www.prolific.com). They were paid $1.6, and the median time to complete the studies ranged between 5.5 and 6.5 minutes. Participation was restricted to English-speaking US-based participants, with an approval rate of at least 95%, who did not perform any of the other tasks in the study (including pilot studies, see below). We did not independently collect participants' age or gender (as we were interested in overall mental simulation capacity limits), and so we report the aggregated results based on the information collected by Prolific, which participants gave consent to use.

Given that Imagined Objects Tracking is a novel task, we ran pilot studies (with the same tasks described in the pre-registration; data available at the OSF) to determine the necessary sample size for both within- and between-subjects comparisons. The smallest effect size found (an interaction of a within-subjects effect and experiment) was partial $\eta^2 = 0.19$, which requires $N = 18$ in each experiment for 95% power with $\alpha = 0.05$ (calculated using G*Power 3,[44]. As a conservative estimate, we decided to double this number in the full study. In the case of participants failing the comprehension questions and being excluded, we recruited additional participants to reach 36. All decisions of screening and re-recruitment were based on pre-registered criteria, and took place without analyzing the data itself.

Participants were excluded from all analyses if they gave an incorrect answer to at least one of the pre-task quiz questions, or (in experiments with 2 entities) if less than 75% of their trials included two unique responses (i.e., two different response keys). To ensure $N = 36$ participants in the final sample of each experiment, this required recruiting $N = 47$, $N = 71$, $N = 53$, $N = 56$, $N = 58$, and $N = 69$ participants in Experiment 1a, 1b, 2, 3, 4, and 5, respectively. This was a comparable rate to similar past studies of intuitive physics conducted online[19,21]. The pre-task quiz and unique responses threshold were the only criteria used for excluding participants, intentionally focusing only on task comprehension rather than task performance. The final sample in Experiment 1a included 16 people who identified as female, 19 as male, and one who preferred not to state (mean age 34.4); Experiment 1b included 18 people who identified as female and 18 as male (mean age 38.6); Experiment 2 included 20 people who identified as female and 16 as male (mean age 40.3); Experiment 3 included 22 people who identified as female, 13 as male, and one who preferred not to state (mean age 37.0); Experiment 4 included 15 people who identified as female and 21 as male (mean age 37.8); Experiment 5 included 24 people who identified as female and 12 as male (mean age 38.2).

Stimuli and procedure

We used an animated dynamic prediction task, similar to tasks previously used to study intuitive physical reasoning[19,21]. Participants in the Imagined Objects Tracking task continue the trajectory of objects in their mind's eye, and a measure of the tracked motion is compared with the ground truth. In the case of the present experiments, the measure is the time in which an event happened in imagination (a collision with the ground), and the ground truth is extracted from the physics engine used to create stimuli. Demos of the tasks are available online: https://jatos.mindprobe.eu/publix/Z8AtkMP8NZt (Experiment 1a), https://jatos.mindprobe.eu/publix/5iB7OvSVmHX (Experiment 1b), https://jatos.mindprobe.eu/publix/1ygkkWoPJZm (Experiment 2), https://jatos.mindprobe.eu/publix/4vh34OQB263 (Experiment 3), https://jatos.mindprobe.eu/publix/Brm7G1m7ogT (Experiment 4), and https://jatos.mindprobe.eu/publix/GX5Rdu8q3VP (Experiment 5).

In all experiments, participants watched short 2D animations, created in the physics engine Pymunk, that used the same simple setting (except for Experiment 5, see below): A green rectangle at the bottom represented the ground, a narrow upright gray rectangle at the horizontal mid-line represented a wall, and a light blue rectangle acted as background. Additionally, each scene included 1 or 2 balls, rendered as yellow or purple disks. For scenes with 2 balls, one was always to the left of the wall and the other was to the right, and the balls differed in color. Scenes started with each ball having some initial height and velocity, after which the balls moved according to simulated physics. We manipulated the initial height and velocity to produce different trajectories that varied in their paths and the time it took a ball to hit the ground, which was either 1.0, 1.2, 1.4, or 1.6 seconds from the start of the animation.

The task was to indicate when a ball touches the ground. Response keys were spatially mapped to avoid confusion: 'F' for balls left of the wall, and 'J' for balls right of the wall (in Experiment 5, balls that move towards the left or right sides of the screen, respectively). No feedback was given following button presses. Each combination of

true impact time and ball movement type was presented 8 times (4 on each side, randomized order), for a total of 64 experimental trials, presented in 2 blocks with a self-timed break between them.

Prior to the test trials, participants went through a pre-task phase, including instructions, practice trials, and a multiple-choice quiz. Each question in the quiz focused on a different aspect of the task (the goal, when animations terminate, how to be most accurate, and response mapping). If a participant failed to respond correctly to any of the 4 quiz questions, they were removed from further analysis. We next provide details specific to the setup of each Experiment.

Experiments 1a and 1b. Animations in the experimental phase paused after 0.5 s (well before either ball touched the ground). Participants were asked to continue the animation in their mind's eye and indicate when the balls in their imagination hit the ground. The starting height and velocity of the balls were chosen such that neither variable on its own determined the time it took for the ball to reach the ground, to discourage the use of heuristics. During practice, participants completed 4 trials with animations that ran all the way through (showing the impact of the ball with the ground), in which they were asked to press a button when they saw the ball touching the ground. This was followed by 4 trials with animations that paused early (not showing the impact), as in the actual experiment. Animations in Experiment 1a showed a single ball (see Fig. 2, left), and animations in Experiment 1b showed 2 balls (see Fig. 3, top left). In two-entity animations, one ball moved in a hyperbole up and to the center, without hitting the wall (from shortest to longest true impact time, these balls started either 100, 140, 60, or 180 pixels above the ground, and their vertical velocity was 95, 110, 216, or 180 pixels per second; their initial distance from the wall was 185 pixels, and their horizontal velocity was 100 pixels per second), and the other ball moved down and towards the wall on a sure collision path with it, but with the moment of collision occurring after the animations paused (from shortest to longest true impact time, these balls started either 280, 480, 400, or 440 pixels above the ground, and their vertical velocity was 190, 200, 60, or 20 pixels per second; their initial distance from the wall was 185 pixels, and their horizontal velocity was 280 pixels per second). The color of the balls was matched to movement type within participants, but randomized between participants. Each movement type was counterbalanced to appear on each side of the wall on half of the trials. Single ball animations were created by removing one of the balls from the 2-balls scenes. In Experiment 1a, each block included only balls either left or right of the wall, with the order counterbalanced across participants.

Experiment 2. The stimuli were identical to Experiment 1b (2 entities), except that the animations continued until participants gave 2 responses, including the moment in which the balls touched the ground, and up to 4 seconds (see Fig. 4, left). So, rather than continue the motion of objects in their imagination, participants were asked to simply click on the appropriate button when they saw the relevant ball touch the ground. Accordingly, the practice phase included 4 full-length animations, without imagination trials.

Experiment 3. The task was identical to Experiment 1b (2 entities, animations pause, participants continue the motion in their imagination). However, unlike Exp 1b, all balls moved in a hyperbole motion, in the same direction (left or right), and velocity was kept constant (see Fig. 5, left). This was designed to create strong motion grouping cues. The only variable that led to different true impact times was the starting height of each ball. The color mapping (yellow/purple ball shown on left/right of wall) was randomized between participants, but constant within participants.

Experiment 4. The task was again the same as in Experiment 1b, and the stimuli were also identical except as noted below. First, the balls did not freeze but continued to move. Second, a gray rectangle occluded much of the bottom part of the scene (see Fig. 6, left). The dimensions of the occluder on each trial were chosen so that the balls disappeared behind it after 500 ms of movement. It spanned vertically from the top of the ground to the bottom point the ball that collides with the wall reached at 500 ms, and horizontally from the side of the video on the colliding ball's side (to hide how it rolls on the ground post-hit) to the innermost point the hyperbole ball reached at 500 ms. Third, the wall was presented in black instead of gray so it is salient against the gray 'screen'. Fourth, the instructions, example trials, and quiz questions were updated to explain the occlusion. Participants first saw unoccluded trials during practice, and then occluded trials.

Experiment 5. The task was similar overall to Experiment 1b: scenes showed two objects pausing mid-motion, and the task was to continue trajectories in the mind's eye and indicate when each object collides with a specific area. The difference was that the animations and description given to participants were altered in the following ways (see Fig. 7, left): We created videos using the same physics engine as before, but with gravity turned off. The scene background was gray, with two black rectangles spanning the height of the video placed on the right and left edges. Two colored disks were presented roughly in the center of the screen, and moved in straight lines and with a constant speed towards the right and left sides of the scene. Participants were asked to press the left side key when the left side disk collides with the left side wall, and the right side key for the right side disk and right side wall.

Analysis

Responses were aggregated across movement type, color, and side. In Experiment 3, all motion paths were hyperbolic, and trials were aggregated across movement direction (towards or away from the center). Individual trials were rejected from further analysis if they did not include two unique responses, as this prevents mapping each response to a specific ball. This rejection was not applied to Experiment 1a, as it involved only one ball in each trial. Trials were also rejected if responses were farther than 3 SDs from a participant's mean. Taken together, these criteria resulted in a rejection of less than a single trial on average in all experiments: 0.3, 0.3, 0.2, 0.5, 0.1, and 0.4 trials on average in Experiment 1a, 1b, 2, 3, 4, and 5 respectively (note that the values reflect the number of rejected trials, not the percentage of rejected trials, which was 0.5%, 0.5%, 0.3%, 0.8%, 0.2%, and 0.6% respectively, all lower than 1% of rejected trials).

Statistical tests. In all experiments, we analyzed Subjective Impact Time using a within-subject Analysis of Variance (ANOVA) with True Impact Time (1.0, 1.2, 1.4, or 1.6 s; extracted from the physics engine) as a factor. In experiments that included 2 entities, we added Response Order (first vs. second key press) as a factor. We followed the ANOVAs with a polynomial contrasts analysis, to test the linear trend of the True Impact Time factor. As a measure of the effect of Response Order in experiments with 2 entities, we used 1000 bootstrap samples to estimate the 95% confidence interval (CI) on the intercept of linear fits, separately for the first and second responses. Our pre-registered predictions were to find (1) a linear trend for all experiments, showing that overall people are sensitive to ground truth physics; (2) a large delay between responses in Experiment 1b, in line with a Serial model; (3) a reduced effect in Experiment 2; (4) an intermediate effect in Experiment 3; and (5) a large serial delay in Experiments 4 and 5. Because the effect of response order was expected to be significant even for Experiment 2 (given that the second response is by definition slower), we additionally compared the effect of Response Order across experiments, using ANOVAs with Response Order as a within-subjects factor, and Experiment as a between-subjects factor, and predicted significant interactions. Supplementary Note 1 further reports a

post-hoc analysis focused on the variation in response times. Violations of sphericity were handled via Greenhouse-Geisser corrections[45]. All tests are two-tailed.

Parallel vs. serial mental simulation models. We created two mental simulation models: Serial and Parallel. Both models relied on the same physics engine that generated the stimuli to simulate the balls, starting from the animation's end point and until both balls collide with the ground. The models differed in how they advanced the objects (see Fig. 1). The Parallel Model moves both balls simultaneously. The Serial Model first picks one ball, advances its state until collision with the ground, then repeats this process for the second ball. More formally, taking a scene to be a tuple of objects o_t^j at time t, and each object to be a list maintaining the properties of the object at time t, and Φ to be the transition function that updates the properties of objects according to physics, we have for the Parallel Model:

$$\begin{aligned} t = 0 &: [o_0^1, o_0^2] = G + \xi, \\ t > 0 &: [o_{t+1}^1, o_{t+1}^2] = [\Phi(o_t^1), \Phi(o_t^2)] \end{aligned} \tag{1}$$

where G is the ground-truth state of the objects as handed by perception, and ξ is the perceptual noise. Following standard practice in modeling intuitive physics with mental simulation[14] we assume this is a two-dimensional Gaussian with mean $\mu = (0, 0)$ and a symmetrical standard deviation $SD = (\sigma^j, \sigma^j)$ for each object j. We estimated an upper level for perceptual noise in an independent experiment (see Supplementary Note 2) and used this value as our noise level, but importantly, our results are robust both below and above this chosen perceptual uncertainty setting, including both no-noise situations, and far greater noise levels (for the full details, see Supplementary Note 3).

For the Serial Model, we have:

$$\begin{aligned} t = 0 &: [o_0^1, o_0^2] = G + \xi, \\ t > 0 \wedge t < C &: [o_{t+1}^1, o_{t+1}^2] = [\Phi(o_t^1), o_t^2], \\ t > C &: [o_{t+1}^1, o_{t+1}^2] = [o_t^1, \Phi(o_t^2)], \end{aligned} \tag{2}$$

where the choice of simulating object o^1 first is arbitrary, and C is the time at which object o^1 collides with the ground. While our main analysis takes the choice of which object to simulate first to be random, we do expect that people are biased in this selection, and indeed we found evidence that people use simple imperfect cues in this selection (see Supplementary Note 1).

Model Fitting. Because the models include perceptual uncertainty, we sampled 20 starting states for each model and averaged the results across runs. In addition to the perceptual uncertainty parameter that was estimated through independent participant data (Experiment S1), we assume that the model response can be fit to the human response up to a simple linear transformation, meaning *Human Subjective Impact Time* $= a \cdot$ (*Model Predicted Impact Time*) $+ b$. We fit the slope and intercept of this linear transformation for each model separately, using the response data of the relevant experiments that involve 2 objects. To compare model performance, we calculated each model's explained variance, as well as the resulting MSE. Model parameter fits were done for the mean responses of all participants. In Supplementary Note 3, we additionally present model fits for individual data (still fitting overall a and b). All of these different analyses agree with the results of the analysis we present in the main text.

Reporting summary

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

Data availability

All of the materials for this study (including raw data minimally analyzed to maintain anonymity, modeling codes, analysis codes, and stimuli) are available in the Open Science Framework: https://doi.org/10.17605/OSF.IO/WZT98.

Code availability

The code for the stimuli, models, and analysis can be found at https://doi.org/10.17605/OSF.IO/WZT98.

References

1. Pylyshyn, Z. W. & Storm, R. W. Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* **3**, 179–197 (1988).
2. Alvarez, G. A. & Franconeri, S. L. How many objects can you track?: evidence for a resource-limited attentive tracking mechanism. *J. Vis.* **7**, 14 (2007).
3. Pylyshyn, Z. W. Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Vis. Cogn.* **11**, 801–822 (2004).
4. Scimeca, J. M. & Franconeri, S. L. Selecting and tracking multiple objects. *WIREs Cogn. Sci.* **6**, 109–118 (2015).
5. Scholl, B. J. What have we learned about attention from multiple-object tracking (and vice versa)? In: Dedrick, D. & Trick, L. (eds.) *Computation, Cognition, and Pylyshyn*, 49–78 https://direct.mit.edu/books/book/3844/chapter/125934/What-Have-We-Learned-about-Attention-from-Multiple (The MIT Press, 2009).
6. Holcombe, A. *Attending to Moving Objects*. 1 ed., https://www.cambridge.org/core/product/identifier/9781009003414/type/element (Cambridge University Press, 2023).
7. Feria, C. S. Speed has an effect on multiple-object tracking independently of the number of close encounters between targets and distractors. *Atten. Percept. Psychophys.* **75**, 53–67 (2013).
8. Franconeri, S., Jonathan, S. & Scimeca, J. Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychol. Sci.* **21**, 920–925 (2010).
9. Lovett, A., Bridewell, W. & Bello, P. Selection enables enhancement: An integrated model of object tracking. *J. Vis.* **19**, 23 (2019).
10. Franconeri, S. L., Pylyshyn, Z. W. & Scholl, B. J. A simple proximity heuristic allows tracking of multiple objects through occlusion. *Atten. Percept. Psychophys.* **74**, 691–702 (2012).
11. Keane, B. & Pylyshyn, Z. Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cogn. Psychol.* **52**, 346–368 (2006).
12. Iordanescu, L., Grabowecky, M. & Suzuki, S. Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *J. Vis.* **9**, 1–1 (2009).
13. Ahuja, A. & Sheinberg, D. L. Behavioral and oculomotor evidence for visual simulation of object movement. *J. Vis.* **19**, 13 (2019).
14. Battaglia, P. W., Hamrick, J. B. & Tenenbaum, J. B. Simulation as an engine of physical scene understanding. *Proc. Natl Acad. Sci. USA* **110**, 18327–18332 (2013).
15. Gerstenberg, T., Goodman, N. D., Lagnado, D. A. & Tenenbaum, J. B. A counterfactual simulation model of causal judgments for physical events. *Psychol. Rev.* **128**, 936–975 (2021).
16. Ullman, T. D., Spelke, E., Battaglia, P. & Tenenbaum, J. B. Mind games: game engines as an architecture for intuitive physics. *Trends Cogn. Sci.* **21**, 649–665 (2017).
17. Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A. & Tenenbaum, J. B. Eye-tracking causality. *Psychol. Sci.* **28**, 1731–1744 (2017).
18. Hamrick, J. B., Battaglia, P. W., Griffiths, T. L. & Tenenbaum, J. B. Inferring mass in complex scenes by mental simulation. *Cognition* **157**, 61–76 (2016).

19. Ludwin-Peery, E., Bramley, N. R., Davis, E. & Gureckis, T. M. Broken physics: a conjunction-fallacy effect in intuitive physical reasoning. *Psychol. Sci.* **31**, 1602–1611 (2020).
20. Ludwin-Peery, E., Bramley, N. R., Davis, E. & Gureckis, T. M. Limits on simulation approaches in intuitive physics. *Cogn. Psychol.* **127**, 101396 (2021).
21. Bass, I., Smith, K. A., Bonawitz, E. & Ullman, T. D. Partial mental simulation explains fallacies in physical reasoning. *Cogn. Neuropsychol.* **38**, 413–424 (2021).
22. Keogh, R. & Pearson, J. The perceptual and phenomenal capacity of mental imagery. *Cognition* **162**, 124–132 (2017).
23. Ceja, C. R. & Franconeri, S. L. Difficulty limits of visual mental imagery. *Cognition* **236**, 105436 (2023).
24. Erlikhman, G., Keane, B. P., Mettler, E., Horowitz, T. S. & Kellman, P. J. Automatic feature-based grouping during multiple object tracking. *J. Exp. Psychol.: Hum. Percept. Perform.* **39**, 1625–1637 (2013).
25. Scholl, B. J. & Pylyshyn, Z. W. Tracking multiple items through occlusion: clues to visual objecthood. *Cogn. Psychol.* **38**, 259–290 (1999).
26. Kosslyn, S. M., Thompson, W. L. & Ganis, G. *The Case for Mental Imagery* (Oxford University Press, 2006).
27. Lau, J. S.-H. & Brady, T. F. Noisy perceptual expectations: Multiple object tracking benefits when objects obey features of realistic physics. *J. Exp. Psychol.: Hum. Percept. Perform.* **46**, 1280–1300 (2020).
28. Oberauer, K. Access to information in working memory: exploring the focus of attention. *J. Exp. Psychol.: Learn., Mem. Cogn.* **28**, 411–421 (2002).
29. Balaban, H., Smith, K. A., Tenenbaum, J. B. & Ullman, T. D. Electrophysiology reveals that intuitive physics guides visual tracking and working memory. *Open Mind* **8**, 1425–1446 (2024).
30. Spivey, M. J. & Geng, J. J. Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychol. Res.* **65**, 235–241 (2001).
31. Pathak, A., Patel, S., Karlinsky, A., Taravati, S. & Welsh, T. N. The "eye" in imagination: the role of eye movements in a reciprocal aiming task. *Behav. Brain Res.* **441**, 114261 (2023).
32. Krasich, K., O'Neill, K. & De Brigard, F. Looking at mental images: eye-tracking mental simulation during retrospective causal judgment. *Cogn. Sci.* **48**, e13426 (2024).
33. Cowan, N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* **24**, 87–114 (2001).
34. Dennett, D. C. *Consciousness Explained* (Penguin UK, 1993).
35. Bear, D. M. et al. Physion: Evaluating physical prediction from vision in humans and machines. preprint at *arXiv* https://arxiv.org/abs/2106.08261 (2021).
36. Vivanco, V., Tenenbaum, J., Paulun, V. C. & Smith, K. Ensemble physics: perceiving the mass of groups of objects is more than the sum of its parts. https://osf.io/jbhzp_v1 (2025).
37. Smith, K. & Vul, E. Sources of uncertainty in intuitive physics. *Top. Cogn. Sci.* **5**, 185–199 (2013).
38. Balaban, H. & Ullman, T. D. Physics versus graphics as an organizing dichotomy in cognition. *Trends Cogn. Sci.* https://doi.org/10.1016/j.tics.2025.05.003 (2025).
39. Makovski, T. & Jiang, Y. V. Feature binding in attentive tracking of distinct objects. *Vis. Cogn.* **17**, 180–194 (2009).
40. Bahrami, B. Object property encoding and change blindness in multiple object tracking. *Vis. Cogn.* **10**, 949–963 (2003).
41. McCloskey, M., Caramazza, A. & Green, B. Curvilinear motion in the absence of external forces: naïve beliefs about the motion of objects. *Science* **210**, 1139–1141 (1980).
42. Li, Y. et al. An approximate representation of objects underlies physical reasoning. *J. Exp. Psychol. Gen.* **152**, 3074–3086 (2023).
43. Peer, E., Brandimarte, L., Samat, S. & Acquisti, A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. *J. Exp. Soc. Psychol.* **70**, 153–163 (2017).
44. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
45. Greenhouse, S. W. & Geisser, S. On methods in the analysis of profile data. *Psychometrika* **24**, 95–112 (1959).

Acknowledgements

We thank members of the CoCoDev labs at Harvard, and the CoCoSci lab at MIT, for their insightful comments. This work was supported by the Center for Brains, Minds, and Machines (HB, TDU), the Defense Advanced Research Projects Agency (DARPA) Machine Common Sense Program (HB, TDU), Israel Science Foundation (ISF) Grant No. 2067/24 (HB), the Alon scholarship (HB), and the Jacobs Foundation (TDU).

Author contributions

Conceptualization: H.B., T.D.U. Methodology: H.B., T.D.U. Formal analysis: H.B. Writing - original draft: H.B. Writing - review & editing: H.B., T.D.U. Visualization: H.B. Funding Acquisition: H.B., T.D.U.

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/s41467-025-61021-8.

Correspondence and requests for materials should be addressed to Halely Balaban.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.