Trends in Cognitive Sciences

CellPress

Opinion

Physics versus graphics as an organizing dichotomy in cognition

Halely Balaban[1] and Tomer D. Ullman[2,*]

People build world models that simulate the dynamics of the real world. They do so in engineered systems for the purposes of scientific understanding or recreation, as well as in intuitive reasoning to predict and explain the environment. On the basis of a major split in the simulation of real-time dynamics in engineered systems, we argue that people's intuitive mental simulation includes a basic split between physical simulation and graphical rendering. We first show how the separation between physics and graphics relies on a natural division of labor in any cognitive system. We then use the physics/graphics distinction to tie together and explain a range of classic and recent findings across different domains in cognitive science and neuroscience, including aphantasia and imagery, different visual streams, and object tracking.

Highlights

Many current cognitive models of mental simulation draw inspiration from game engines – engineering tools that are used to create and transform scenes for animations and video games in an approximate and efficient way.

A central division of computational labor in modern game engines differentiates between physical simulation and graphical rendering, and this may have a cognitive equivalent.

Physical simulation is central to cognition broadly, and may support most types of mental imagery as well as object tracking.

Aphantasia, the apparent lack of voluntary imagery, is a topic of intense current focus in cognitive science and philosophy, and there are debates about whether it indicates a true lack of visual representations; it may be or can be explained by 'broken rendering'.

Simulations in the mind and in machines

It is a classic view in cognitive science that people use mental simulation to understand the world [1,2]. In recent years, this proposal has been the target of major renewed focus. It has acquired new layers and new bruisings, specifically with regards to one of the pillars of common sense – intuitive physical reasoning (e.g., [3–6]).

In the decades during which mental simulation was debated in cognitive science, other areas of research developed better and better simulations of reality. For present purposes, the most relevant advances in simulation have been in computer graphics. These simulation tools form the backbone of modern games and animations. They are concerned with constructing and manipulating realistic-enough worlds, under heavy constraints of time, memory, and computation.

The computational constraints on approximate world simulations make engineered tools a model system for how mental simulations work cognitively and neurally [5,7–9]. But a foundational principle of engineered simulations has not yet been turned into a central coherent principle in the mental simulation proposal. This is the basic dichotomy between physical simulation and visual rendering – between geometry and graphics. We argue here for a cognitive version of this division. If true, this split would form a major organizing principle in the human brain, with separate processes for physical simulation and graphical rendering.

In what follows, we first briefly present the idea of engineered game engines and their application to cognitive science. We next describe specifically the proposed dichotomy between physical simulation and graphical rendering. We then show how this distinction applies to major current debates and sheds light on a diverse set of important recent findings and puzzles across mental imagery, neuroscience, and object tracking.

[1]Department of Education and Psychology, The Open University of Israel, Raanana 4353701, Israel
[2]Department of Psychology, Harvard University, Cambridge, MA 02138, USA

*Correspondence: tullman@fas.harvard.edu (T.D. Ullman).

Background: game engines and intuitive physics

Game engines are a set of specialized interacting software modules that support the development of animations and games [10,11]. The particular modules and how they move, draw, and show entities vary, but they are all concerned with creating scenes, moving entities using dynamics, and displaying how those entities look using graphics. Most simulation engines contain a 'physics engine' that is dedicated to updating a scene and running it forward. To save on computational resources, physics engines do not fully solve the equations that govern the equivalent scene in reality. Instead, they rely on simplified representations of objects and dynamics, as well as on specialized but principled approximations, shortcuts, and workarounds.

The efficient and approximate nature of physics engines has drawn the attention of many cognitive researchers who study how humans intuitively reason about, predict, and explain the behavior of everyday objects. This intuitive physics is a major part of commonsense reasoning [12], is early-developing or possibly innate, and is likely shared across cultures and with non-human animals [7,13]. Although people are adept at reasoning about and interacting with everyday dynamics [3], empirical work has also shown that people make systematic mistakes in physical reasoning [14,15], and this poses a major concern in education and pedagogy of formal physics [16].

In the past decade some researchers suggested that intuitive physical reasoning relies on a mental game engine simulation. The simulation approach has old roots [17,18], but the flourish of game engines shaped this idea into new forms, with a specific emphasis on how physics engines produce good-enough solutions via principled workarounds [5] (Box 1). In recent years, this approach has been used to explain reasoning in a wide variety of contexts, including trajectories and collisions [19], causality and counterfactuals [6], stability [3], fluids [20], tool-use [21], and cognitive development [22]. The 'physics engine in the head' has also found support in recent neuroscience work [23,24].

Simulation engines are not the only proposal for the computations that underlie our intuitive physics. Alternatives include impetus theory [16,25], heuristics [26], logical rules [27], qualitative reasoning [28], bottom-up features [29], and slot-based deep learning [30]. The debates regarding intuitive physics are important and ongoing, but it is not our purpose here to adjudicate between studies that emphasize successes versus failures of intuitive physics, or to argue for the very

Box 1. Approximations and limitations in mental simulation

Engineers looking to create adequate or 'good enough' simulations often face limitations in computational resources. Similarly, the mind is under resource limitations in general [101] and in mental simulation more specifically [3,5]. Through convergent conceptual evolution, the same approximations and principled work-arounds that are useful in engineered simulation engines may be used by the mind. Such approximations create systematic deviations from a 'perfect' simulation. Different approximations apply to the 'physics' and 'graphics' aspects of the computational pipeline, although many of the approximations studied in recent years pertain to physical simulation and include (i) rough 'bodies' that do not perfectly overlap with fine-grained shapes, leading to systematic deviations in tracking and causal reasoning [32]; (ii) partial simulation, such that only some objects are dynamically moved in a scene, leading to systematic deviations in judgments [31]; (iii) capacity limits in the number of objects simulated in parallel [102]; (iv) finite resources that make a simulation more coarse past a certain point [103]; and (v) lazy evaluation that does not bother setting the properties of objects unless and until required, leading to mental images that contain 'holes' and non-commitment [104]. Many other game engine approximations provide a rich empirical testing ground, such as the distinction between 'dynamic' objects that can be affected by forces in a simulation, and 'static' objects that form the backdrop of a scene and which do not need to be updated, or the distinction between 'awake' objects that require a recalculation of state at every cycle, and 'asleep' objects are not updated unless participating in a collision or change of constraint. Again, such limitations apply independently of whether a scene is 'rendered', and separate considerations or speed-ups apply to the operation of graphical rendering. For example, on the speed-up side, many graphical rendering operations can be carried out in parallel, using specialized architecture. On the limitation side, finite resources can lead to lower resolution, rendering of smaller parts of the scene, and longer 'refresh' rates between rendering cycles.

existence of mental simulation in intuitive physics (discussed in recent related papers; e.g., [4,15,19,31]). We take it that there is sufficient evidence to suggest that people perform reasonably well in many tasks that involve everyday dynamics, that some intuitive physics relies on mental simulation, and that game engines provide a decent framework for capturing mental physical simulation. This is our starting point, and from it we move to the main argument.

The separation between physics and graphics

To recap: intuitive physics is an important part of everyday reasoning. Game engines are frameworks that support approximate simulations. Given similar needs and constraints, game engines are a good candidate framework for intuitive physics. This approach has found empirical, neural, and computational support. Different game engines differ in implementation, but at a high level they use the same principled approximations with possible analogs in the human mind.

Beyond specific workarounds and tailored approximations that help to make a simulation feasible, game engines make a fundamental, ontological separation between physics and graphics – between simulation and rendering. To get a sense of this separation, imagine that an engineer is asked to create a simple game in which players throw an apple at a pile of oranges (Figure 1, Key figure; top). Instead of a scientifically accurate simulation (or a hand-drawn animation), the engineer first constructs a physical scene, placing approximate geometric models of the oranges, apple, floor, etc. The scene can be inspected using visualization tools, but the underlying representation is amodal. In this geometric-spatial representation, there is barely a difference between apples and oranges.

Having set the physical scene, the engineer needs to take a picture of it. This rendering creates the graphical display of how the scene appears to a camera located in a specific location (Figure 1, top right). Rendering means solving graphical computations related to shading and lighting, and requires finer-grained representations of objects. Because graphical and physical computations are different, many entities in game engines lead a dual life [32] – they have a physical representation that carries the information necessary for simulation (position, weight, elasticity, orientation, rough extent, etc.), and a graphical representation necessary for rendering (including color, gloss, texture, and exact extent).

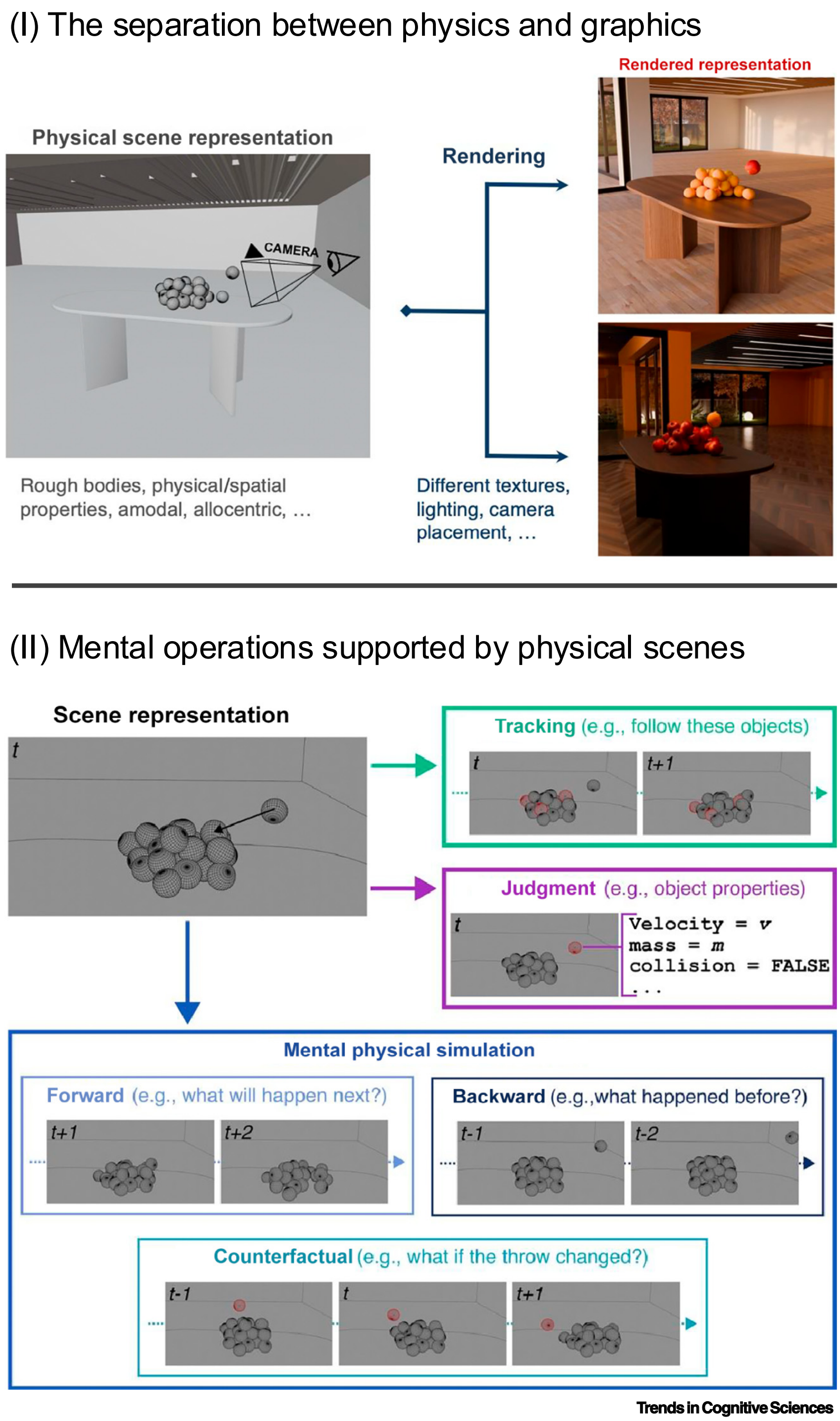
Two crucial issues to note are that physical modeling and graphical rendering in engineered systems can be independent, and that for many purposes rendering is costly and unnecessary. For example, changing the colors and textures of objects would change the rendering of the scene without affecting the physical simulation. If the game changed to throwing oranges at apples rather than apples at oranges, the underlying physical scene would stay the same, but the game would need to change textures and rerender (Figure 1, top right).

We argue that the human mind also has a division of labor between physics and graphics. These are handled by distinct cognitive processes, with distinct types of representations, and distinct neural architectures. Such a division is useful *a priori*: for the purposes of recognition and selection (distinguishing apples from oranges), 'graphical' visual attributes are crucial. But, for the purposes of tracking and action (throwing or tracking apples), surface-level visual attributes do not matter, whereas rough shape, weight, velocity, and location do.

Beyond theoretical motivations, we next consider how the dichotomy between physics and graphics can apply to a variety of findings across cognitive science and neuroscience. To preempt a specific objection, we note that 'related-but-not-the-same' divisions – between object identity and space, or between features-for-recognition and features-for-action – have been proposed in

Key figure

Physical scene representations



(See figure legend at the bottom of the next page.)

related fields before. We unpack the similarities and differences with these previous proposals as they become relevant.

The physics and graphics division applied to cognition

The separation we have drawn between physics and graphics can apply either bottom-up (perception, scene interpretation; e.g., [33]) or top-down (imagery/imagination). We next consider several outcomes of a physics–graphics distinction to important, related, yet distinct areas in cognition: the imagery debate, the origins of aphantasia, the ventral–dorsal pathways taxonomy, and object tracking.

Simulation versus rendering in imagery and aphantasia

There have been many sides to the vast, ongoing, decades-long discussion over the format of mental imagery ('The Imagery Debate'). Two major camps include those who argue that mental images are propositional, (e.g., [34], and those who argue that mental images are analog and picture-like (e.g., [18,35]). Mental imagery goes beyond visual imagery, and includes auditory, gustatory, tactile, and action-based imagery. Researchers have also proposed the notion of 'spatial imagery' as distinct from other forms of imagery, with distinct neural (e.g., [18,36,37]) and behavioral [38–40] signatures. While details differ by framework regarding what spatial imagery captures as compared to visual or object-based imagery, it is generally seen as referring to the locations of objects in space, and to abstract spatial relations between objects or object parts.

Our view is that the 'physics' part of scene construction maps onto, explains, and expands the notion of spatial imagery, whereas the 'rendering' operation corresponds to the creation of images in sensory areas, equivalent to the depiction of an image from a particular perspective. As mentioned, previous research has proposed that 'spatial' mental imagery is functionally distinct from 'object-based' and 'visual' imagery. However, such spatial imagery is intended to capture information about distances between objects and their parts, as well as their relative locations, such that it is a distinct operation to imagine 'object O' (visual) or 'in location L' (spatial). This is not our proposal. Instead, 'spatial imagery' should be expanded to cover 'physical imagery', including all information necessary for reasoning about motion and interaction. Physical-spatial imagery encompasses the entire 'scene', including many object properties, including extent. Physical imagery is fundamentally object-based, but the objects it traffics in are the geometric representations necessary to simulate a scene and its dynamic evolution (Figure 1). Note that, although physical simulation does not involve rendering, it can store non-physical object features (e.g., material), just as an engineer's amodal scene representation may note that an apple has 'red' in its list of properties but without rendering it.

Figure 1. (I) Schematic of the separation between physics and graphics. Paralleling the separation that exists in modern simulation engines, we suggest that there is a principled cognitive split between the physical scene representation and the rendered graphical representation which rely on different properties and computations. The physical scene representation includes rough object extents, locations, and properties needed for simulation and tracking. On the basis of this representation, a graphical 'rendering' operation can create a pixel-based representation showing what the scene looks like from a particular point of view, and different graphical choices could lead to different renders while keeping the physical/spatial scene the same. Note that the physical scene representation is amodal and allocentric (in simulation engines as well as in our cognitive proposal); the use of a specific point of view and (gray) colors for the physical scene in this figure is strictly to get the point across. (II) Physical scene representations support different operations. The physics engine-like non-graphical depiction (top left) consists of a coarse object-based representation of a scene, with approximate bodies, locations, forces, and so on. Much of mental imagery (bottom) can be performed via mental physical simulation without rendering: forward for prediction, backwards for inference, and conditioned on changed variables to support counterfactual reasoning. Other operations include tagging different objects in the perceptual scene to support tracking (top right), and querying the properties of objects (e.g., mass) to support judgment (middle right).

The proposal to divide imagery by physics and graphics aligns in part, and importantly diverges from, the two main views on mental imagery. In our view, most visual imagery tasks are solved through physical simulation, including scanning, rotation, navigation, and manipulation. The situation is similar to how engineers do not need to 'render' a scene to simulate it or answer many questions about it. Although the pipeline can render images in a pixel-like format (which can serve as an input for further bottom-up processing), this rendering and reprocessing is not strictly necessary for most tasks. The rendered output itself is mostly epiphenomenal. In this sense, the propositional camp was right [41]. However, rendering in our proposal involves scene representations that are the functional equivalent of objects with spatial extent, similar to the analogical representation of a mesh. In this sense, the depictive camp was right [35].

It is helpful to see how the distinction explains a specific example. Consider mental rotation, in which comparing images takes longer as the angle between the depicted entities grows [2], as if people perform stepwise rotation of (depictive) mental images.

However, propositional accounts point out that mental rotation is cognitively penetrable, is influenced by task demands [41], and is not reinterpreted: when meaningful shapes appear in atypical orientations, mentally rotating them to a typical orientation does not lead to recognition [42]. We suggest that mental rotation is done via physical imagery: internally simulating geometric transformations of 'wire-frame' analogical entities, creating strong spatial effects (including activation of visual-spatial brain areas due to differential attendance to areas of the scene [43]) alongside difficulties in tracking surface features during rotation [44]. Graphical rendering is unnecessary, and if used can be performed solely for the final physical/spatial representation rather than continuously [45].

The distinction between rendering and simulation in mental imagery plays out most clearly in explaining cases where people seemingly do not create visual images at all, but can solve visual imagery tasks. This is the case of 'aphantasia' – the inability to willfully create visual sensations in the mind's eye. First noted more than a century ago [46], in recent years aphantasia has become a major topic of research [47–49]. While the initial reliance on subjective reports led some to suggest that 'aphantasia' reflects disagreement about terminology, further research showed people with aphantasia differ from the general population in ways suggesting that they truly lack imagery [50–53].

Given that non-subjective measures across different tasks suggest that aphantasia truly reflects a lack of mental imagery, it is surprising how little this lack seems to affect other abilities. Broadly, people with aphantasia function much like people without it [54], including in tasks where a widely reported strategy in the general population is mental imagery [52,55]. This is underscored by the fact that the phenomena of aphantasia drew little attention until the past decade or so, and many people with aphantasia report being unaware until a relatively late age that their inner life is different from others. People with and without aphantasia also score similarly overall on tasks involving spatial imagery or memory reconstruction of the spatial (but not visual) details of images [56,57], as well as mental rotation [58]. For mental rotation, there are some small differences in timing [59], but these may be the result of the non-simulation stages.

The pattern of results in aphantasia poses a puzzle. The similarities to the general population strongly suggest that people with aphantasia can solve tasks that were typically thought to involve mental imagery. The differences strongly suggest that people with aphantasia are not simply misunderstanding the terminology, but truly lack the ability to create mental images. There have been at least two major and opposing proposals for how to resolve this conundrum [60]. The first proposal accepts the lack of mental imagery, and suggests that the success of people with aphantasia is proof of the existence of multiple strategies for solving various tasks (e.g., [59]): people without aphantasia rotate objects in their mind using mental imagery, whereas people with aphantasia do ... something else (Figure 2, left). The alternative proposal accepts the successes, but insists that the only way to solve such tasks is through mental imagery. The subjective experience of people with aphantasia and the non-subjective differences they show on some tasks reflects intact visual mental imagery, but a lack of higher-level access to these representations [61,62]. On this proposal, people with aphantasia actually do create functional visual images to rotate objects in the mind, but they are simply unaware of it (Figure 2, middle).

We have a different suggestion, based on the dichotomy between geometry and graphics: aphantasia is a 'broken' rendering operation. The situation is akin to an engineer who can create a physical model of a scene, but cannot then use rendering to create a pixel-based representation showing the scene from a particular perspective, with early visual areas playing the role of the canvas (note that this does not mean that the entire graphical 'module' is affected; e.g., [63]). In our view, people with aphantasia truly do not have visual imagery representations in the sense of rendered images, and this explains why their top-down activation in early visual areas is weaker and less perception-like [64–66]. However, for most tasks that supposedly involve mental imagery, they are using the same representations and strategies that people without aphantasia are using (Figure 2, right). This view dovetails with the recent proposal of 'blindsight in imagery' [67]. But, while that proposal can be taken to mean aphantasia involves images being rendered without awareness, for us, people with aphantasia do not have visual imagery that they are

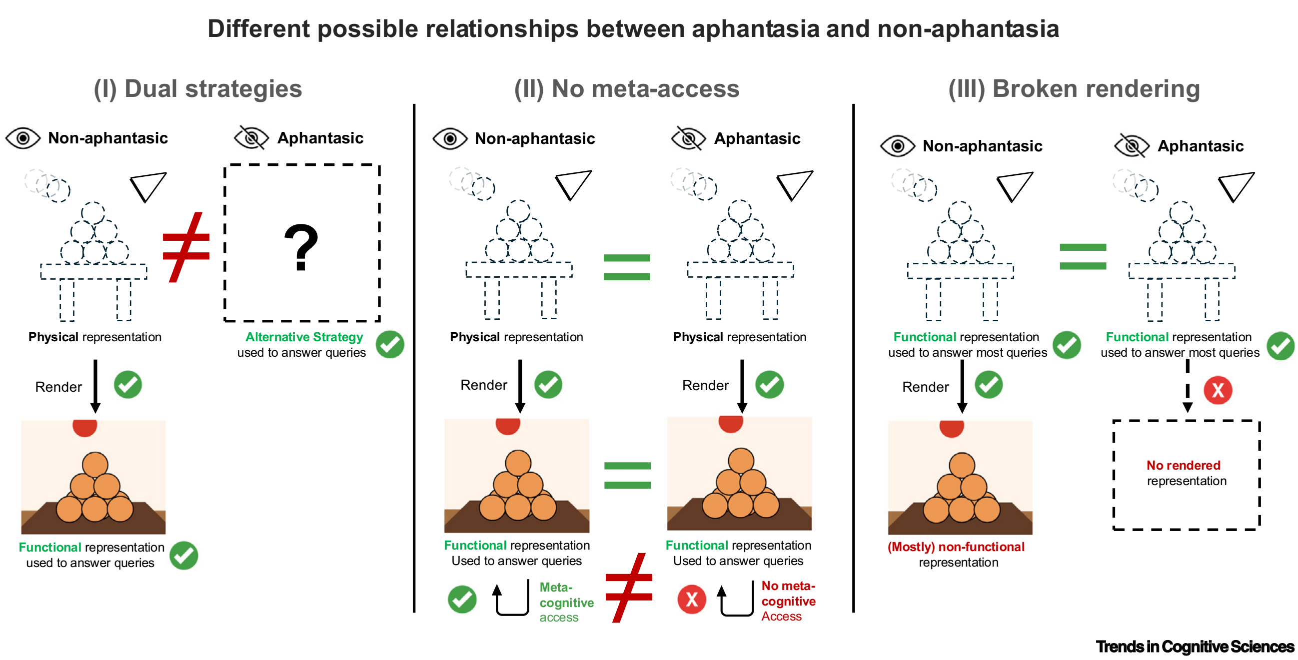


Figure 2. Schematic summary of several different views of aphantasia. (I) Dual strategies suggest that people without aphantasia use rendered visual mental images to solve most tasks associated with subjective visual imagery, such as mental rotation and scanning, whereas people with aphantasia have developed non-visual image-based strategies that happen to mostly line up with them [56]. (II) No mental access suggests that both people with and without aphantasia rely on rendered visual images to solve various tasks, but people with aphantasia are simply not meta-cognitively aware of their own visual images [58]. (III) Broken rendering, the view we endorse here, suggests that people with aphantasia do not create visual mental images, but such representations are not functional in most tasks for people without aphantasia, and that the two populations share the underlying, relevant, non-visually experienced physical representation that is used to solve most imagery tasks.

unaware of. Instead, people with intact visual imagery solve those tasks not via visual imagery but via physical imagery, just like people with aphantasia.

Our view explains how aphantasia is primarily a lack of top-down visual imagery, and can also account for the few (but important) cases in which there are measurable non-subjective differences between people with and without aphantasia. These include the lack of internally generated attentional templates [68] or priming effects [50], and less-efficient attentional guidance in visual search [69], which should rely on the top-down activation of graphically detailed representations. It can also account for the finding that fear responses in people with aphantasia are limited to perceptually presented visual stimuli and missing for verbal descriptions ([51], also [70]): modules in the mind that are attuned to perceptual stimuli need such low-level visual features as input to trigger a response. Being unable to (top-down) activate the 'pixel-like' canvas means that modules which rely on it for input do not trigger a response. Similarly, other differences between people with aphantasia and controls (e.g., [71]) likely point to tasks aided by visual imagery in the latter. However, the successes in most situations and tasks that are taken to require visual mental imagery should cause us to realize that the functional 'imagery' part of these tasks is physical/spatial. For most of the imagery tasks that matter and can be tested, the amodal geometric scene representation is sufficient. The subjective visual experience is due to an epiphenomenal rendering of this scene. In many of the ways that matter, we all have aphantasia.

Graphics in the ventral stream, physics in the dorsal

Neuroscience has strongly established that the brain has at least two largely distinct pathways for processing visual information – the dorsal and ventral streams. Patient studies, lesions, and neuroimaging have demonstrated that the ventral (occipitotemporal) pathway is involved in object perception, whereas the dorsal (occipitoparietal) pathway is involved in other types of visually guided behavior. Originally, researchers classified the streams as categorization versus localization, nicknamed 'what versus where' [72,73]. This was refined given evidence of dorsal involvement in action preparation and object manipulation (e.g., [74,75]), leading to terms such as 'what versus how' and 'vision for perception versus vision for action'. The exact definition of the two pathways is a matter of important open debate (e.g., [76,77]).

Applying our game engine dichotomy to brain data, we suggest that the distinction should be between a graphical (ventral) pathway and a geometrical/physical (dorsal) pathway. The role of bottom-up processing in the ventral stream is mostly the same, and subserves functions such as object recognition. This mechanism can also be activated top-down for visual imagery or detailed attentional templates ('rendering'). The more significant amendment is to the dorsal stream, which we suggest is the physics pathway that is in charge of mentally simulating the physics of objects. As with many generative models (Figure 1), this processing can be run 'forward' (top-down) by simulating likely future outcomes from the current state to support action selection and execution, as well as prediction (e.g., [3]). The model can also be run 'backward', going from the observed kinematics to inferred dynamics [78], and these two directions can also be combined to support causal reasoning (e.g., [6]).

As mentioned, a long tradition describes the dorsal stream as spatial. We accept that spatial information has a central role in occipitoparietal processing, but argue that this is due to the relevance of spatial information for physics [79]. In this view, the ventral stream also 'cares' about spatial information, specifically the details necessary for graphical rendering and de-rendering from particular points of view [80], and the dorsal stream also 'cares' about object representations, in particular coarse body approximations that are separate from fine-grained structure [81,82]. Our proposal does not mean that all spatial reasoning is physics-based. Although many tasks currently described as 'spatial' may rely on a mental physics engine, some tasks might be accomplished via other routes. This would explain the mixed findings regarding the connection between physical and spatial abilities [83–85], but more work is necessary to directly test how physical and spatial reasoning relate.

Our suggested division of labor aligns with recent empirical work, specifically fMRI findings of a frontoparietal brain network implicated in physical reasoning, representing physical properties, and unfolding scenes [8,9,23,79] (Box 2). The same frontoparietal areas are also activated when people engage with tools or plan grasp actions [86,87] which led researchers to describe them as "the brain's first-person physics engine" [8], in line with the current proposal that the dorsal stream operates a cognitive physical simulation process.

Object tracking relies on physics, not just on spatiotemporal information

Researchers have long used [88], and continue to use [89], object tracking as a window into human perception, attention, and cognition. A key tool in this research line is the 'multiple object tracking' task in which people must keep track of initially marked targets from identical distractors throughout motion. People can only track a few objects, depending on the speed and proximity of the entities [90–92]. This is intuitive. What is less intuitive is that, even for the few items they do track, people struggle to report their features – people are better at indicating whether an item was a target than which specific target it was [93–96]. Adding to the puzzle, only some features seem to directly matter for tracking. Empirical research separates spatiotemporal and non-spatiotemporal features, but it is not clear why this separation exists given that it is not locations *per se* that are being tracked [97,98].

Instead of spatiotemporal separation, we suggest a physics-based approach (Figure 1). Tracking the trajectory of moving items can be accomplished by an approximate physical simulation of their future locations (potentially rotating between targets on a single-item serial basis [89]). A

Box 2. Insights from injuries

Our proposal leads us to expect a dissociation between physics and graphics following specific injuries or neurodegeneration. Although research to directly examine this is a topic for future research, several findings align with the proposal.

One line of support for specialized processing of physical information comes from recent findings on semantic dementia [105], a loss of semantic memory that is associated with cortical atrophy in the temporal lobe. Despite a broad deficit across verbal and non-verbal semantic tasks, physical knowledge appears to be spared in these patients, who perform normally in tests of tool-use, which are in turn strongly related to physical understanding [106].

In different but related work, patients with ventral-stream injuries show impairments in both visual recognition and visual imagery, but intact performance in types of reaching and grasping [107]. Although this has been interpreted as a separation between 'object' and 'spatial' processing, people's ability to interact with objects in such aphasias goes beyond a general sense of location and seems to include a rough understanding of objects, corresponding perhaps to the notion of an approximate body or convex hull in a physics engine.

Several dorsal-stream deficits also map naturally onto physically relevant information, most obviously motion blindness in akinetopsia [108], movement coordination deficits in optic ataxia [109], and problems in executing voluntary movements as well as tool-use in apraxia [110–112]. Two other syndromes, hemispatial neglect (a lack of awareness or attention to the side contralateral to the brain damage) and simultanagnosia (an inability to perceive more than a single object at a time) might reflect a problem in generating or maintaining the scene representation that is central to mental physical simulation.

In very recent work, it was found that a patient with hemispheric neglect could nevertheless report on various features of the 'neglected' side of an object [113]. For example, when seeing a chimeric object made by fusing a red swan (left) and blue truck (right), the patient reports that it is a half-red, half-blue truck (correctly processing color but incorrectly ignoring the chimeric nature of the object). Such behavior may reflect a neglect of shape, but an interesting possible interpretation would be that purely graphical attributes such as color are 'filled in' within a shape representation that is faulty (either because the rough convex shape is not fully available or because it cannot be integrated with fine shape details), similarly to in-painting in current image-generation algorithms.

physics-based prediction mechanism for tracking also naturally accounts for the factors that influence tracking performance: velocity is one of the most important features for simulation, and how close items are to each other is especially relevant for collision detection. Further, a physical tracking system would mean that surface features are not what people are using for tracking. So, it is expected that they fail to notice when these features change, and also fail to report them even for items that were successfully tracked. Physics engines also represent objects via coarse body approximations –, an approach which people also adopt even in paradigms that are not traditionally linked with intuitive physics, such as change detection [32], and which can be used to successfully model human performance in tracking and reasoning tasks [22].

Our suggestion is then that object tracking relies on core physical knowledge more generally rather than on spatiotemporal continuity specifically (as has been previously suggested; e.g., [99]). Beyond behavioral evidence, this is in line with recent electroencephalography (EEG) findings [100] showing that violations of intuitive physics disrupt object tracking, but tracking is restored when such violations are explained away, even for events that were identical in all spatiotemporal information and only differed in their high-level physical explanation.

Concluding remarks

The distinction between physical simulation and graphical rendering is the backbone of engineered systems that aim to mimic the dynamics of the world in real-time. Such a distinction is also likely to be a basic organizing principle in mental processing. On the basis of this distinction, we suggest that much of the functional use of mental imagery is driven by physical simulation rather than by graphical rendering, that aphantasia reflects 'broken rendering', that different visual streams in the brain map onto the graphics–physics distinction, and that classic and recent puzzles in object tracking can be explained on the basis of this organizational principle. The foundational split between physics and graphics is the starting point for much additional research, and some directions of particular interest are outlined in the Outstanding questions. To highlight just one: if rendering is mostly epiphenomenal, why bother? Why waste time and energy putting paint on a canvas that does not matter? While the point is to raise the question rather than answer it, we note this would not be the first time an artist did something just for the hell of it.

Outstanding questions

What is the 'point' of rendering, if any? Physical simulation is sufficient to solve many tasks taken to involve mental imagery, and although fine-grained attentional templates may benefit from rendering, this is a small benefit for such a costly operation.

How do the physical and graphical systems interact? In engineered simulation engines the physical/spatial scene is primary and object-centered, whereas graphical rendering is secondary to it, but merging graphical and physical information presents a new type of binding problem for cognition.

How do social modules interact with physics and graphics? Nearly all engineered simulation engines have agents in them which are both physical objects as well as sources of decisions and actions beyond a physical simulation, but it is unclear whether agent representations in the mind should be seen as part of the physical simulation or are separate from it.

What are the developmental trajectories for physics and graphics? Given the primary of spatial/physical simulation in games, it may develop earlier than graphical rendering, and this option remains open.

What is the evolutionary trajectory for the physics–graphics split? Because many of the tasks that seem to rely on this split are shared with non-human animals, it seems unlikely to be unique to humans, but further comparative work will be necessary to show how and when such a split emerged.

How do dreams map onto the physics–graphics distinction, if at all? Although they are notoriously challenging to empirically study, dreams are widely reported to involve sensory experiences (even among people with aphantasia), but could it be that functional aspects of dreams are better seen as being related to simulation?

Acknowledgments

We would like to thank Matan Mazor and members of the Computation, Cognition, and Development Laboratory at Harvard for important suggestions and useful discussions. Figure 1 is based on renderings by Dragoljub Nikolov. This work was supported by the Center for Brains, Minds, and Machines (H.B., T.D.U.), the Defense Advanced Research Projects Agency Machine Common Sense Program (H.B., T.D.U.), the Alon Scholarship (H.B.), the Israel Science Foundation (grant 2067/24 to H.B.), and the Jacobs Foundation (T.D.U.).

Declaration of interests

The authors declare no competing interests.

References

1. Craik, K.J.W. (1943) *The Nature of Explanation*, Cambridge University Press
2. Shepard, R.N. and Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science* 171, 701–703
3. Battaglia, P.W. *et al.* (2013) Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18327–18332
4. Ludwin-Peery, E. *et al.* (2021) Limits on simulation approaches in intuitive physics. *Cogn. Psychol.* 127, 101396
5. Ullman, T.D. *et al.* (2017) Mind games: game engines as an architecture for intuitive physics. *Trends Cogn. Sci.* 21, 649–665
6. Gerstenberg, T. *et al.* (2021) A counterfactual simulation model of causal judgments for physical events. *Psychol. Rev.* 128, 936–975
7. Spelke, E.S. (2022) *What Babies Know: Core Knowledge and Composition*, Vol. 1. Oxford University Press
8. 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
9. Pramod, R.T. *et al.* (2022) Invariant representation of physical stability in the human brain. *Elife* 11, e71736
10. Gregory, J. (2018) *Game Engine Architecture*, CRC Press
11. Smith, A.R. (2021) *A Biography of the Pixel*, MIT Press
12. Ullman, T.D. and Tenenbaum, J.B. (2020) Bayesian models of conceptual development: Learning as building models of the world. *Annu. Rev. Dev. Psychol.* 2, 533–558

13. Lin, Y. *et al.* (2022) Infants' physical reasoning and the cognitive architecture that supports it. In *The Cambridge Handbook of Cognitive Development* (Houdé, O. and Borst, G., eds), pp. 168–194, Cambridge University Press
14. McCloskey, M. *et al.* (1980) Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. *Science* 210, 1139–1141
15. Ludwin-Peery, E. *et al.* (2020) Broken physics: a conjunction-fallacy effect in intuitive physical reasoning. *Psychol. Sci.* 31, 1602–1611
16. Hubbard, T.L. (2022) The possibility of an impetus heuristic. *Psychon. Bull. Rev.* 29, 2015–2033
17. Hegarty, M. (2004) Mechanical reasoning by mental simulation. *Trends Cogn. Sci.* 8, 280–285
18. Kosslyn, S.M. *et al.* (2006) *The Case for Mental Imagery*, Oxford University Press
19. Smith, K. *et al.* (2023) Integrating heuristic and simulation-based reasoning in intuitive physics. *PsyArXiv*, Published online July 5, 2023. https://doi.org/10.31234/osf.io/bckes
20. Bates, C.J. *et al.* (2019) Modeling human intuitions about liquid flow with particle-based simulation. *PLoS Comput. Biol.* 15, e1007210
21. Allen, K.R. *et al.* (2020) Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *Proc. Natl. Acad. Sci.* 117, 29302–29310
22. Smith, K. *et al.* (2019) Modeling expectation violation in intuitive physics with coarse probabilistic object representations. *Adv. Neural Inf. Proces. Syst.* 32, 8983–8993
23. Schwettmann, S. *et al.* (2019) Invariant representations of mass in the human brain. *eLife* 8, e46619
24. Fischer, J. (2021) The building blocks of intuitive physics in the mind and brain. *Cogn. Neuropsychol.* 38, 409–412
25. McCloskey, M. and Kohl, D. (1983) Naive physics: the curvilinear impetus principle and its role in interactions with moving objects. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 146–156
26. Proffitt, D.R. *et al.* (1990) Understanding wheel dynamics. *Cogn. Psychol.* 22, 342–373
27. Davis, E. (1990) *Representations of Commonsense Knowledge*, Morgan Kaufmann
28. Forbus, K.D. (2014) Qualitative reasoning about space and motion. In *Mental Models* (Gentner, D. and Stevens, A.L., eds), pp. 61–82, Psychology Press
29. Lerer, A. *et al.* (2016) Learning physical intuition of block towers by example. *Proc. Machine Learn. Res.* 48, 430–438
30. Piloto, L.S. *et al.* (2022) Intuitive physics learning in a deep-learning model inspired by developmental psychology. *Nat. Hum. Behav.* 6, 1257–1267
31. Bass, I. *et al.* (2021) Partial mental simulation explains fallacies in physical reasoning. *Cogn. Neuropsychol.* 38, 413–424
32. Li, Y. *et al.* (2023) An approximate representation of objects underlies physical reasoning. *J. Exp. Psychol. Gen.* 152, 3074–3086
33. Aglioti, S. *et al.* (1995) Size–contrast illusions deceive the eye but not the hand. *Curr. Biol.* 5, 679–685
34. Pylyshyn, Z.W. (2002) Mental imagery: in search of a theory. *Behav. Brain Sci.* 25, 157–238
35. Pearson, J. (2019) The human imagination: the cognitive neuroscience of visual mental imagery. *Nat. Rev. Neurosci.* 20, 624–634
36. Farah, J. (1989) The neural basis of mental imagery. *Trends Neurosci.* 12, 395–399
37. Borst, G. and Kosslyn, S.M. (2010) Individual differences in spatial mental imagery. *Q. J. Exp. Psychol.* 63, 2031–2050
38. Marks, D.F. (1973) Visual imagery differences in the recall of pictures. *Br. J. Psychol.* 64, 17–24
39. Reisberg, D. *et al.* (2003) Intuitions and introspections about imagery: the role of imagery experience in shaping an investigator's theoretical views. *Appl. Cogn. Psychol.* 17, 147–160
40. Blajenkova, O. *et al.* (2006) Object-spatial imagery: a new self-report imagery questionnaire. *Appl. Cogn. Psychol.* 20, 239–263
41. Pylyshyn, Z.W. (1979) The rate of 'mental rotation' of images: a test of a holistic analogue hypothesis. *Mem. Cogn.* 7, 19–28
42. Slezak, P. (1991) Can images be rotated and inspected? A test of the pictorial medium theory. In *Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 13)*, pp. 55–60, Erlbaum
43. Dijkstra, N. *et al.* (2017) Distinct top-down and bottom-up brain connectivity during visual perception and imagery. *Sci. Rep.* 7, 5677
44. Xu, Y. and Franconeri, S.L. (2015) Capacity for visual features in mental rotation. *Psychol. Sci.* 26, 1241–1251
45. Ankaoua, M. and Luria, R. (2022) One turn at a time: behavioral and ERP evidence for two types of rotations in the classical mental rotation task. *Psychophysiology* 60, e14213
46. Galton, F. (1880) Statistics of mental imagery. *Mind* 5, 301–318
47. Greenberg, D.L. and Knowlton, B.J. (2014) The role of visual imagery in autobiographical memory. *Mem. Cogn.* 42, 922–934
48. Zeman, A. *et al.* (2015) Lives without imagery – congenital aphantasia. *Cortex* 73, 378–380
49. Zeman, A. *et al.* (2020) Phantasia – the psychological significance of lifelong visual imagery vividness extremes. *Cortex* 130, 426–440
50. Keogh, R. and Pearson, J. (2018) The blind mind: no sensory visual imagery in aphantasia. *Cortex* 105, 53–60
51. Wicken, M. *et al.* (2021) The critical role of mental imagery in human emotion: insights from fear-based imagery and aphantasia. *Proc. Soc. Biol. Sci.* 288, 20210267
52. Keogh, R. *et al.* (2021) Visual working memory in aphantasia: retained accuracy and capacity with a different strategy. *Cortex* 143, 237–253
53. Milton, F. *et al.* (2021) Behavioral and neural signatures of visual imagery vividness extremes: aphantasia versus hyperphantasia. *Cereb. Cortex Commun.* 2, tgab035
54. Monzel, M. *et al.* (2023) No general pathological significance of aphantasia: an evaluation based on criteria for mental disorders. *Scand. J. Psychol.* 64, 314–324
55. Zhao, B. *et al.* (2022) Spatial transformation in mental rotation tasks in aphantasia. *Psychon. Bull. Rev.* 29, 2096–2107
56. Bainbridge, W.A. *et al.* (2021) Quantifying aphantasia through drawing: those without visual imagery show deficits in object but not spatial memory. *Cortex* 135, 159–172
57. Pounder, Z. *et al.* (2022) Only minimal differences between individuals with congenital aphantasia and those with typical imagery on neuropsychological tasks that involve imagery. *Cortex* 148, 180–192
58. Zeman, A.Z. *et al.* (2010) Loss of imagery phenomenology with intact visuo-spatial task performance: a case of 'blind imagination'. *Neuropsychologia* 48, 145–155
59. Kay, L. *et al.* (2024) Slower but more accurate mental rotation performance in aphantasia linked to differences in cognitive strategies. *Conscious. Cogn.* 121, 103694
60. Zeman, A. (2024) Aphantasia and hyperphantasia: exploring imagery vividness extremes. *Trends Cogn. Sci.* 28, 467–480
61. Liu, J. *et al.* (2023) Visual mental imagery in typical imagers and in aphantasia: a millimeter-scale 7-T fMRI study. *Cortex* 185, 113–132
62. Dijkstra, N. (2024) Uncovering the role of the early visual cortex in visual mental imagery. *Vision* 8, 29
63. Bouyer, L.N. and Arnold, D.H. (2024) Deep aphantasia: a visual brain with minimal influence from priors or inhibitory feedback? *Front. Psychol.* 15, 1374349
64. Cabbai, G. *et al.* (2024) Sensory representations in primary visual cortex are not sufficient for subjective imagery. *Curr. Biol.* 34, 5073–5082
65. Montabes de la Cruz, B.M. *et al.* (2024) Decoding sound content in the early visual cortex of aphantasic participants. *Curr. Biol.* 34, 5083–5089
66. Chang, S. *et al.* (2025) Imageless imagery in aphantasia revealed by early visual cortex decoding. *Curr. Biol.* 35, 591–599
67. Michel, M. *et al.* (2025) Aphantasia as imagery blindsight. *Trends Cogn. Sci.* 29, 8–9
68. Keogh, R. and Pearson, J. (2021) Attention driven phantom vision: measuring the sensory strength of attentional templates and their relation to visual mental imagery and aphantasia. *Philos. Trans. R. Soc. B Biol. Sci.* 376, 20190688
69. Monzel, M. *et al.* (2021) Imagine, and you will find – lack of attentional guidance through visual imagery in aphantasics. *Atten. Percept. Psychophysiol.* 83, 2486–2497
70. Speed, L.J. *et al.* (2024) The role of visual imagery in story reading: evidence from aphantasia. *Conscious. Cogn.* 118, 103645

71. Dance, C.J. *et al.* (2023) The role of visual imagery in face recognition and the construction of facial composites – evidence from aphantasia. *Cortex* 167, 318–334
72. Schneider, G.E. (1969) Two visual systems: brain mechanisms for localization and discrimination are dissociated by tectal and cortical lesions. *Science* 163, 895–902
73. Goodale, M.A. and Milner, A.D. (1992) Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25
74. Almeida, J. *et al.* (2010) The role of the dorsal visual processing stream in tool identification. *Psychol. Sci.* 21, 772–778
75. Shmuelof, L. and Zohary, E. (2005) Dissociation between ventral and dorsal fMRI activation during object and action recognition. *Neuron* 47, 457–470
76. Freud, E. *et al.* (2020) What does dorsal cortex contribute to perception? *Open Mind* 4, 40–56
77. Mahon, B.Z. (2023) Higher order visual object representations: a functional analysis of their role in perception and action. In *APA Handbook of Neuropsychology (Vol. 2: Neuroscience and Neuromethods)* (Brown, G.G. *et al.*, eds), pp. 113–138, American Psychological Association
78. Hamrick, J.B. *et al.* (2016) Inferring mass in complex scenes by mental simulation. *Cognition* 157, 61–76
79. Navarro-Cebrián, A. and Fischer, J. (2022) Precise functional connections between the dorsal anterior cingulate cortex and areas recruited for physical inference. *Eur. J. Neurosci.* 56, 3660–3673
80. James, T.W. *et al.* (2002) Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. *Neuron* 35, 793–801
81. Ayzenberg, V. and Behrmann, M. (2022) Does the brain's ventral visual pathway compute object shape? *Trends Cogn. Sci.* 26, 1119–1132
82. Freud, E. *et al.* (2018) More than action: the dorsal pathway contributes to the perception of 3-D structure. *J. Cogn. Neurosci.* 30, 1047–1058
83. Hart, Y. *et al.* (2018) The statistical shape of geometric reasoning. *Sci. Rep.* 8, 12906
84. Mitko, A. and Fischer, J. (2020) When it all falls down: the relationship between intuitive physics and spatial cognition. *Cogn. Res. Princ. Implic.* 5, 24
85. Mitko, A. *et al.* (2024) A dedicated mental resource for intuitive physics. *iScience* 27, 108607
86. Chao, L.L. and Martin, A. (2000) Representation of manipulable man-made objects in the dorsal stream. *NeuroImage* 12, 478–484
87. Culham, J.C. *et al.* (2003) Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Exp. Brain Res.* 153, 180–189
88. Pylyshyn, Z.W. and Storm, R.W. (1988) Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* 3, 179–197
89. Holcombe, A. (2023) *Attending to Moving Objects*, Cambridge University Press
90. Alvarez, G.A. and Franconeri, S.L. (2007) How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *J. Vis.* 7, 14
91. Intriligator, J. and Cavanagh, P. (2001) The spatial resolution of visual attention. *Cogn. Psychol.* 43, 171–216
92. Holcombe, A.O. *et al.* (2014) Object tracking: absence of long-range spatial interference supports resource theories. *J. Vis.* 14, 1
93. Pylyshyn, Z.W. (2004) Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Vis. Cogn.* 11, 801–822
94. Horowitz, T.S. *et al.* (2007) Tracking unique objects. *Percept. Psychophys.* 69, 172–184
95. Wu, C.-C. and Wolfe, J.M. (2018) Comparing eye movements during position tracking and identity tracking: no evidence for separate systems. *Atten. Percept. Psychophysiol.* 80, 453–460
96. Saiki, J. and Holcombe, A.O. (2012) Blindness to a simultaneous change of all elements in a scene, unless there is a change in summary statistics. *J. Vis.* 12, 2
97. Scholl, B.J. (2001) Objects and attention: the state of the art. *Cognition* 80, 1–46
98. Maechler, M.R. *et al.* (2021) Attentional tracking takes place over perceived rather than veridical positions. *Atten. Percept. Psychophysiol.* 83, 1455–1462
99. Flombaum, J.I. *et al.* (2009) Spatiotemporal priority as a fundamental principle of object persistence. In *The Origins of Object Knowledge* (Hood, B.M. and Santos, L.R., eds), pp. 135–164, Oxford University Press
100. Balaban, H. *et al.* (2024) Electrophysiology reveals that intuitive physics guides visual tracking and working memory. *Open Mind* 8, 1425–1446
101. Lieder, F. and Griffiths, T.L. (2020) Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.* 43, e1
102. Balaban, H. and Ullman, T.D. (2024) The capacity limits of mental simulation. *PsyArXiv*, Published online April 14, 2024. https://doi.org/10.31234/osf.io/xzcmb
103. Wang, Y., and Ullman, T.D. Resource bounds on mental simulations: evidence from a liquid-reasoning task, J. Exp. Psychol. Gen. (in press)
104. Bigelow, E.J. *et al.* (2023) Non-commitment in mental imagery. *Cognition* 238, 105498
105. Chapman, C.A. *et al.* (2020) Evaluating the distinction between semantic knowledge and semantic access: evidence from semantic dementia and comprehension-impaired stroke aphasia. *Psychon. Bull. Rev.* 27, 607–639
106. Baumard, J. *et al.* (2021) Physical understanding in neurodegenerative diseases. *Cogn. Neuropsychol.* 38, 490–514
107. Spagna, A. *et al.* (2021) Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: a meta-analysis of neuroimaging evidence. *Neurosci. Biobehav. Rev.* 122, 201–217
108. Van Swol, J.M. *et al.* (2024) Akinetopsia: a systematic review. *J. Neuroophthalmol.* 44, e483–e488
109. Rossetti, Y. *et al.* (2019) Definition: optic ataxia. *Cortex* 121, 481
110. Stoll, S. *et al.* (2025) Apraxia: from neuroanatomical pathways to clinical manifestations. *Curr. Neurol. Neurosci. Rep.* 25, 1
111. Goldenberg, G. and Hagmann, S. (1998) Tool use and mechanical problem solving in apraxia. *Neuropsychologia* 36, 581–589
112. Goldenberg, G. and Spatt, J. (2009) The neural basis of tool use. *Brain* 132, 1645–1655
113. Karakose-Akbiyik, S. *et al.* (2024) Preserved recognition of basic visual features despite lack of awareness of shape: evidence from a case of neglect. *Cortex* 176, 62–76