Mind Games: Game Engines as an Architecture for Intuitive Physics

Tomer D. Ullman[a,b], **Elizabeth Spelke**[b], **Peter Battaglia**[c], **Joshua B. Tenenbaum***[a]

[a] *Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*
[b] *Department of Psychology, Harvard University*
[c] *DeepMind*

We explore the hypothesis that many intuitive physical inferences are based on a mental physics engine, analogous in many ways to the machine physics engines used in building interactive video games. We describe the key features of game physics engines along with their parallels in human mental representation, focusing especially on the intuitive physics of young infants where the hypothesis helps to unify many classic and otherwise puzzling phenomena, and may provide the basis for a computational account of how infants' physical knowledge develops. This hypothesis also explains a number of "physics illusions", and helps to inform the development of AI systems with more human-like common sense.

Simulating Physics in a Mind and a Computer

Human perception cares about 'what is where', but also 'where to', 'how' and 'why'. We implicitly but continually reason about the stability, strength, friction and weight of objects around us, to predict how things might move, sag, push and tumble as we act on them. As naive observers, people may be most aware of the cases where we get these predictions wrong, but for cognitive scientists seeking to understand how humans interact so flexibly with everyday objects and with each other, or for artificial intelligence (AI) researchers who want to build human-like common sense in machines, what is most striking is how right we are. Even young children have a remarkable capacity for intuitive physics, extending even to objects they are encountering for the first time, yet we are still far from having robots or other AI systems with the physical scene understanding abilities of a human baby, let alone an adult.

Our goal in this paper is to suggest one route for closing this gap, for explaining in engineering terms the core intuitive physics that arises in young children and develops into adulthood, which could also support building these capacities in machines. We call this hypothesis the "game engine in your head": Evolution could equip infants with something like the high-level architecture used to interactively simulate the physics of virtual worlds in modern video games (Fig. 1), and learning physics would then consist of "programming" this architecture

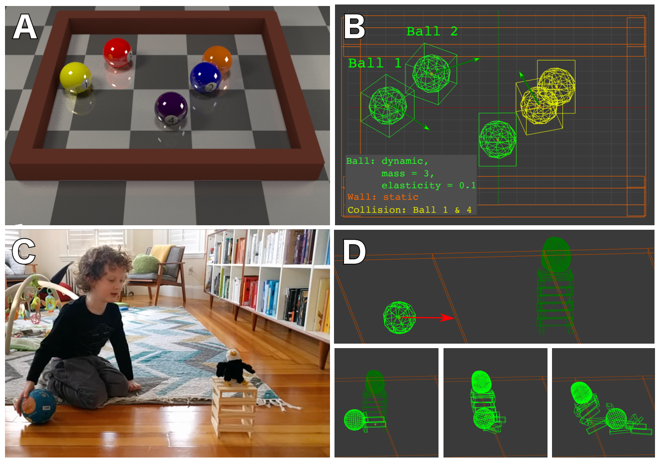

Figure 1: **How game engines view the world.** Everyday perception is not just about the categorization of objects, but about their dynamic properties and relations **(A)** Simple dynamic image, billiard balls colliding in a constrained environment **(B)** Physics engine view of the billiard scene, parsing the world into objects with physical properties, velocity vectors, and events such as collisions **(C)** Prediction in a daily scene **(D, top)** Physics engine representation includes static floor (orange), simplified bounding bodies, force vectors (red arrow), sleeping and waking objects (dark and light green) **(D, bottom)** Simulating forward from initial conditions. The sleeping bodies wake up as the collision moves through the tower.

to better capture the infant's experiences observing and interacting with objects and other physical entities. Compared to other approaches to physical simulation, game physics engines are optimized for efficiency on a limited subset of everyday physics, and for producing results that look natural regardless of their quantitative correspondence to physical reality. Integrated with tools from probabilistic inference and machine learning, game physics-style representations can explain how people are able to make a wide range of intuitive physical judgments quickly and robustly (Box 1), and to acquire many kinds of physical knowledge from experience – including the physics of the world we actually live in, but also possible worlds that humans could experience.

In the following, we introduce the key features of game physics engines that make them compelling models for the representations of intuitive physics, with an emphasis on how these features correspond to important distinctions and developmental milestones that have been discovered in the earliest emerging core intuitive physics of infants (Fig. 2). These infant findings are fascinating but often puzzling, lacking a unifying ex-

*Corresponding author
Email address: jbt@mit.edu (**Joshua B. Tenenbaum***)

planation. Strikingly, game physics engines predict many of these results, and perhaps can provide the missing integrative theory for infant physics, while also being consistent with the adult cognitive state and plausible learning mechanisms (Box 2). We also show how physics engine concepts can make sense of several kinds of "physics illusions" that people are prone to, as byproducts of the short cuts they make for efficient simulation, and discuss how they are also being used and extended by AI researchers to build more human-like physical reasoning and planning in machines. Finally, we briefly discuss ways in which people's intuitive physics may differ from or go beyond what game engines naturally represent (Box 3). We do not mean to suggest that all the inner workings of physics engines will have counterparts in the mind, or that people's understanding of all aspects of the physical world depends on a mental physics engine. The parallels suggest, however, that the mental physics engine hypothesis provides insights into diverse aspects of human physical reasoning and especially its developmental origins.

Major Physics Engine Concepts

Our proposal can be seen as one computational instantiation of the classic view that intuitive physics is enabled by "runnable mental models": mental simulators that to a certain degree capture the causal mechanisms at work in the world [1], and can be evolved forward to predict and reason about objects' dynamics mechanically and spatially [e.g. 2]. The extent to which human intuitive physics in fact relies on something like a simulation engine, and in what situations is the engine applied, are open questions subject to ongoing debate [3, 4]. Here we will take as a starting point that simulation provides a powerful mechanism for at least some intuitive physical inferences, and present game physics engines as a candidate computational substrate for those simulations.

Mental models and mental simulation processes in intuitive physics have often been seen as qualitative in nature, with a mathematical basis that is fundamentally different from scientific mechanics [5]. Recently, however, the notion of a mental physics simulator that supports quantitative inferences has led to strong computational models for a wide range of intuitive physical judgments, physical scene understanding, and counterfactual reasoning [e.g. 6, 7, 74, 9, 10, 11, and see Box 1]. These models combine advances in probabilistic reasoning in AI with the exciting technological developments in physics engines that have taken computer animation from block shapes to blockbuster movies and games. The video game industry in particular has developed tools for building rich, immersive environments that must react convincingly and in real time to the open-ended actions of players exploring them from a first-person perspective.

We believe that these same tools provide a first working hypothesis for the representational contents of intuitive physics: the data structures that our minds use to represent the objects and events that make up a scene, and the algorithms we use to simulate physical dynamics over time.

There are several reasons why game physics engines may be useful representations for cognitive scientists to explore. The first is that game physics engines are programmed by humans, for humans. Their functionality may thus provide hypotheses for what passes as a "good enough" approximation to real-world physics, as humans understand it. But we expect there may be deeper analogs between the computational architecture of physics engines in video games, and the mental architecture that lets humans grasp and predict the immediate future of physical scenes, because both of these systems evolved under similar design constraints and pressures: Neither is required to capture physics exactly or perfectly; both were designed to produce reasonable-looking dynamic approximations of complex scenes on a human-relevant scale, in real time, with computational resources far too limited to implement anything like a precise molecular simulation. Contrast this for example with scientific physical simulations of galaxy formation, atomic systems, weather patterns, or protein folding. Such simulations have been essential tools in scientific research, but in these settings, simulations can draw on vast computational resources and take much longer than real time; there is no reason to think that their fundamental representations parallel any concepts we expect to be relevant for everyday human cognition.

Crucially, while game physics engines share some of the quantitative structure of Newtonian mechanics or classical fluid mechanics, they also depart dramatically from these scientific models, both in how they represent the world, and how they are used to reason about the world. It is very unlikely that the human mind solves the equations of motion that fully describe a complex dynamical scene, as physicists do when they compute trajectories over long time intervals, such as the orbits of the planets, or the arc of a cannonball. Game physics engines do not carry out these computations either. Rather, they use a combination of approximations to Newtonian mechanics that are highly computationally efficient when run forward one step at a time; hacks and shortcuts that have no scientific basis but produce plausible dynamics very efficiently; and qualitative switches between different approximation schemes at salient points in space and time [12]. Indeed, for both game engines and the brain's simulations, the models don't have to be accurate in any sense that physicists would recognize; they just have to produce results that look reasonable at the spatial scales that humans perceive and act on, and predict well enough over a short time interval of a second or two. To be useful, they have to make these predictions fast – faster than real time. They have to be flexible, to handle a very large number of situations, including quite novel ones. And they have to run on low power circuitry – a brain, or a smart phone.

To give a high-level example of how physics engine approximations work, and the kinds of trade-offs they make, consider different ways we could simulate the relatively simple scene of several billiard balls moving and colliding in a closed space (Figure 1A and B). One option is to create a simulation that is veridical as possible, down to the molecular level. Such a simulation might be the most physically accurate, but it is far too computationally-intensive for real-time applications. Another option, pursued by many neural network models, is to treat the scene as a single high-dimensional vector in some latent representational space [e.g. 13], undergoing a complex non-linear evolution, and attempt to directly predict the next state of this vector given the current one and past statistical regularities.

A physics engine represents a middle way between these extremes: instead of 10^{26} particles, or a single high-dimensional vector, the engine explicitly divides the world into a relatively small number of individuated objects that occupy space, with properties that may be stable or changing in time (billiard ball, table, wall, mass, friction, position, and so on). This factorization into objects, just as in Newtonian mechanics, abstracts and simplifies the scene to enable efficient computation. But the actual computations in physics engines hack Newtonian mechanics in many ways. For instance, nearly all physics engines separate objects' dynamics into free-motion and collision-solving phases. As long as the objects are not colliding with other objects or surfaces, they move roughly according to $F = m \cdot a$, within constraints. A separate collision-detection module spots when objects overlap, and switches their dynamics into a collision-resolution mode. Often, this collision-detection module does not take into account the specifics of the object's shape, and instead uses a simplifying bounding box to notice overlaps. Furthermore, the simulation can usually assume that many entities in the scene (such as walls, floors or background objects) are not in motion, and thus require no moment-to-moment computations to update their position.

In the rest of this section, we consider these and other "physics engine hacks" in more detail, with a focus on how they parallel core phenomena in perception and cognition, and especially how they provide insight into the object and event representations of young infants.

Objects and Events Two foundational representational commitments of game physics engines are objects and events. Objects are bounded chunks of matter in space, events are delimiting points in time and the periods between them. For example, in order to render a scene of several balls colliding, for example, a physics engine explicitly represents these balls as named entities with location and velocity, size and shape, mass and elasticity, and so on. When the balls overlap in space, this triggers a specific collision event in the physics engine, which alters the dynamics of the objects (and see the segment on Collision Detection). This may seem such an obvious representation that one can well ask how it could possibly be otherwise, but recent work on building artificial systems with a sense of intuitive physics has focused instead on representing a physical scene as a vector of pixels, without an explicit notion of objects or events [see 13, and see the Section on Intuitive Physics in AI, Machine Learning, and the Brain].

According to several proposals, infants from early in their development also see scenes as made up of objects and events [e.g. 14, 15, 16, 17]. Infants group parts of a scene into holistic entities based on their motion, and have certain physical expectations about these entities: they should continue existing, not suddenly change direction, not interpenetrate, and so on. Infants are also sensitive to subtle qualitative differences in the events that describe the motion of these objects, distinguishing between collision, occlusion, stability, containment, and so on.

Box 1. Physical Scene Understanding

The mental physics-engine hypothesis proposes that people reason about physical scenes in the following way: First, people reconstruct the visible scene internally, with some uncertainty over the perceptual and physical properties of the objects (e.g. position, velocity, mass, and friction). This reconstruction is similar to that of a software engineer who looks at a tower of blocks on a table and re-creates an approximation of the objects and dynamics on her computer's physics engine. People can mentally interact with this scene and simulate its future state, repeatedly and with noisy Newtonian dynamics. Such a mental simulation is similar to a set of repeated computer simulations of a tower of blocks by a software engineer, who can predict how a tower of blocks will fall if it is bumped into, using her computer-simulated tower (see Figure I). People can also compare the predictions of the simulation with observations, and adjust their beliefs accordingly. Think of an engineer who wrongly predicts a tower of blocks will collapse when jostled because her computer simulation predicted a collapse, and who readjusts the physical parameters (e.g. the mass of the blocks) of her simulation accordingly (and see Figure 4).

Mental physics engines support a variety of predictions across different tasks and types of reasoning, including predicting how a scene will unfold over time [6, 74], interacting with a dynamic scene [75], reasoning about underlying physical properties [10, 11, 66], causal judgments [7], and quantitative infant physical reasoning [9].

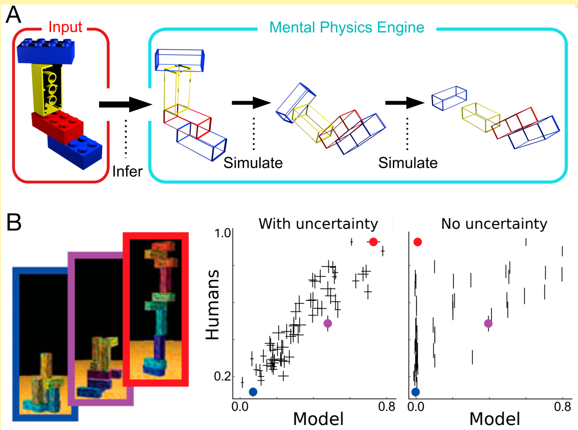

Figure I. Predicting Stability. A mental physics engine model vs. human judgment averages for judging the stability of towers. (A) A physics engine simulates the dynamics of inferred towers (B) Each point in the correlation graphs represents people's stability judgments for one tower (with SEM), and the three colored circles correspond to the three towers shown on the left. Ground-truth physics (no uncertainty) does not correspond to human judgments, but a noisy physics simulation does. Adapted from [6].

Static and Dynamic A common way to save on computation time and memory is to classify entities into those that actively participate in the simulation (dynamic or active), and those that do not (static or passive).

Static entities often form the background to a scene, such as walls or the ground. Static structures are not just large-mass objects, they form a separate ontological category, often with zero or undefined mass, for which forces and various other updates are not calculated. Dynamic objects are not simply entities currently in motion, but rather those with the potential to be affected by forces. This basic distinction between static and dynamic could also hold in mental physics engines, from early in development and onwards, explaining how infants and adults come to have different expectations about the physics of static and dynamic entities, about the likely behavior of balls vs. walls.

This distinction is in keeping with various findings, among them the fact that extended surfaces are used early on in navigation (while everyday objects are not), explained by the expectation that such extended surfaces are stable and unlikely to move and therefore reliably indicate one's position [18, 19]. This expectation of stability and immobility is also used for body orientation, shown by the shift in posture and loss of balance in both adults and young children when perceiving a moving three-sided room [See 20, 21, and Figure 2A]. The viewers in these experiments assume the walls of the room are static, and incorrectly infer from their apparent motion that they themselves must be falling.

Beyond orientation and navigation, this distinction can explain certain object groupings and motion predictions, such as why 3-month-old infants expect heterogeneous objects to be grouped and moved together regardless of discontinuities in color and shape, but do not group those objects with the stage floor on which the objects stand. For example, infants who see a hand lifting the top of an object made of two distinct parts expect the entire structure to rise regardless of the discontinuity, but do not expect the floor to come with it [14]. They treat the floor as an immovable, static background, whereas the object itself is dynamic.

Sleeping and Awake Within the category of dynamic objects, physics engines treat objects at rest and objects in motion differently. There is no need to calculate equations of motion for objects that are not in motion. Also, there is usually no need to re-render an object (i.e., re-draw fully its graphical counterpart) if it did not move since the last frame. Objects in a state of rest are labeled 'sleeping'. A sleeping object wakes up if a body collides with it, or if one of its supports (another object or joint) is moved or destroyed. An awake object is put to sleep if its velocity remains below some ϵ threshold over a period of simulation steps S.

For mental physics engines, the concept of a 'sleeping' object can also reduce cognitive load on attention and computational resource allocation. In a typical scene, most (non-agent) entities are not moving at any given time, even though they can potentially be moved given the right force application. The categorical distinction between sleeping and waking entities can account for key findings in the psychology of causality. Consider a rolling billiard ball A hitting a stationary ball B and sending it rolling. People often see this event as A causing B to move, rather than B causing A to stop or slow down [22, 23, 24]. Infants respond to reversals of such events as indicating a change in causal roles [25]. From a purely Newtonian physics perspective, A and B are on equal footing. From a physics-engine perspective, however, the order of events is as follows (Figure 2B):

1. Awake body A moving towards sleeping body B.
2. Collision detected.
3. The status of object B changes to 'awake'.
4. Collision resolved, new velocities assigned.
5. Simulation resumes.
6. Optional: A's new velocity is below threshold ϵ. After several simulation steps S, the engine sets the velocity of A to 0 and puts it to sleep.

Step 2+3 indicate a change of state for B, and directly relate it to A's contact with B. The change of state for A, if it happens, occurs several simulation steps after the collision, and is not directly related to the collision. This basic asymmetry in the state change of the physics engine is in line with the apparent causal asymmetry.

The sleep/wake divide can shed light on findings showing piecemeal mechanical simulation in adults [2]. When asked to predict the behavior of a mechanical system such as an arrangement of gears or pulleys, adults often answer as though they mentally animate pieces of the scene separately, propagating effects through a causal chain rather than simulating the whole scene holistically. Such a causal propagation can be seen in a physics engine as the effect of one moving object waking up the other objects it encounters through collision or force.

Beyond questions of causality and changes of state, the sleep/wake divide is connected to the greater degree of attention people pay to moving objects, and to the role played by motion in the assignment of object boundaries. Presented with a stationary array of novel objects whose boundaries are not clear, people who attempt to move things around will perceive two things that move together as lying on the same object, whereas two things that are too heavy to budge will continue to have indeterminate status. These perceptions are shared by infants in the first months of life [26, 27] and even newborns [28], who use common states of motion - but not common states of rest - as a cue for determining object grouping and boundaries (Figure 2C). From a purely Newtonian perspective a stationary center-occluded object is just as unitary as a uniformly moving one: both have the same motion vector above and below the occluder. Not so from a physics-engine perspective, where objects that are not in motion can be temporarily omitted from the simulation. If, on top of predicting the motion of objects, a mental physics engine is tasked with reconstructing object identities from perceptual data, it would save on computation and memory to have as few moving items to track as possible. Two nearby perceptual patches with identical velocity vectors would be more efficiently characterized as one object.

Detecting and Resolving Collisions All game and physics engines that move objects around must notice when those objects interact, and adjust their motion appropriately. These computations are usually handled by a specialized collision-detection module, though simulators use a variety of methods to detect and solve collisions. To detect collisions, some simulators advance the simulation by a small step and create a list of the overlapping bodies, while other simulators cast trajectories geometrically into the future and check for intersections.

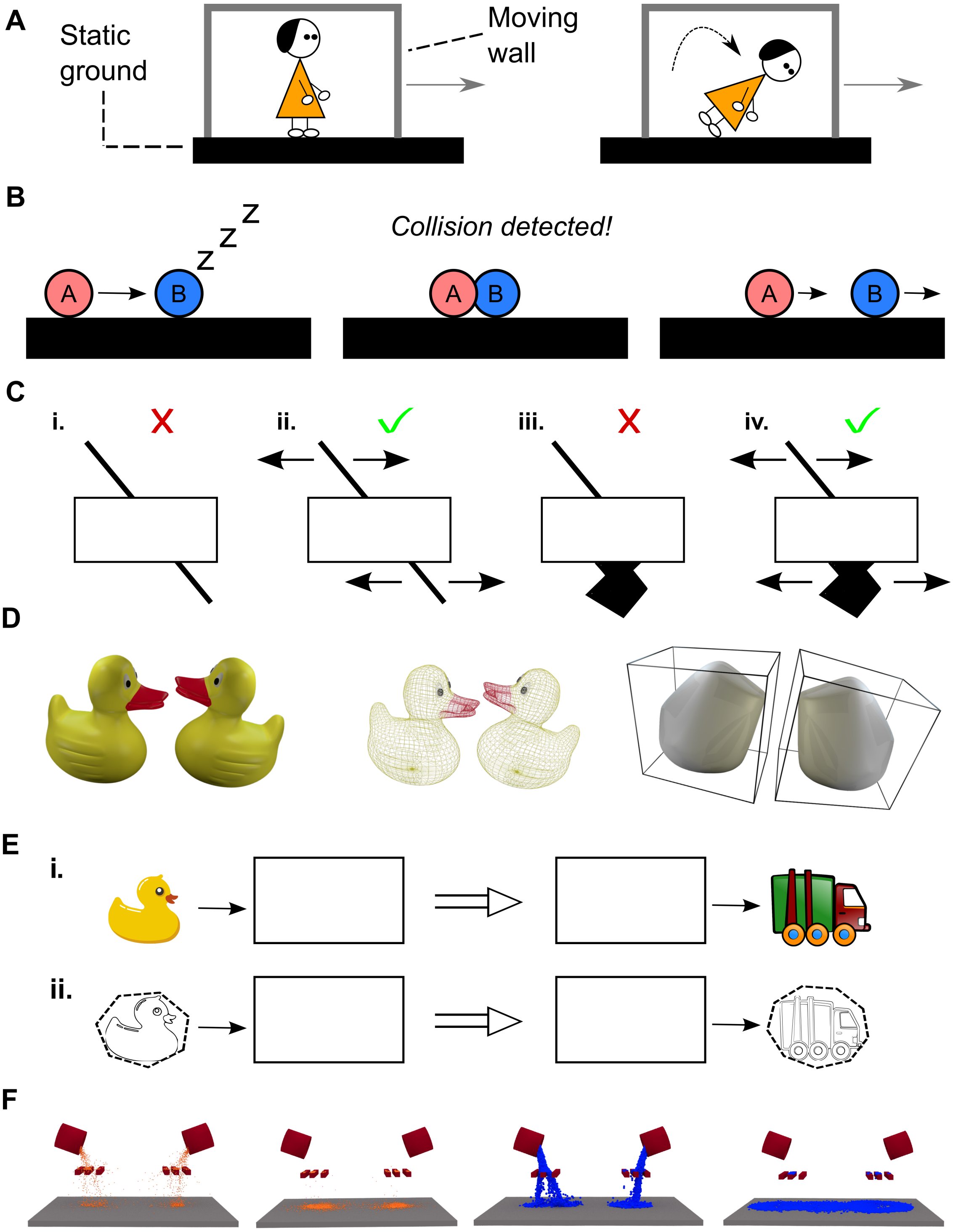

Figure 2: **Experimental findings and computational proposals.** **(A)** Adults and children expect a static structure such as a wall not to move, and its motion is interpreted as self-motion, leading to needless correction and imbalance [20]. **(B)** In a physics engine, the resting body is in a sleep state to save on computation. Following a collision, the body is woken up. **(C)** (i-iv) Young infants use motion, not continuation cues, to perceive connected objects behind an occluder. Green and red marks indicate when infants perceived an occluded object as a unified body (adapted from [26]) **(D)** Game engines distinguish between the visual shape and the related body of an object. The graphical shape is ultimately what is rendered on the screen, using e.g. polygon meshes and textures. The physical body is used under-the-hood for quickly determining overlap and applying forces, making use of bounding boxes and convex hulls, for example **(E)** When 10-month-old infants see a duck go behind an occluder and a truck comes out, they do not expect the duck to remain behind the occluder (i). This may be because the duck/truck body-representation is similar (ii). **(E)** Young children have separate expectations for solids and non-solid substances [46, 47], predicting that non-cohesive substances will go around solids, and through porous barriers, for example. Not all substances are the same. For example, sand (orange substance, left) may accumulate in piles, while water (blue substance, right) spreads. A game engine can simulate non-solid substances with different dynamic properties (such as viscosity), to predict different possible outcomes.

To solve collisions, some simulators place springs between the colliding objects, while others simply dictate changes to the object positions ('pushing' them apart) until the objects no longer intersect.

If mental physics engines exist, they will also need to detect and solve collisions. As collision detection is a specific and separate module in nearly all physics engines, we can expect to find high sensitivity to collisions in humans, regardless of specific object identity. Young infants are particularly sensitive to spatio-temporal boundaries in collision detection. They expect solid objects not to inter-penetrate [15], reason about the location, shape and compressibility of an object behind a rotating screen to predict its collision with the screen [29], anticipate that the size of a colliding object will affect how far an object is displaced [30], and expect collisions with inert objects to result by way of direct contact [31].

The mental physics-engine proposal posits that humans are not perfect in their dynamic simulations for several reasons, including perceptual uncertainty (e.g., where is the object), property uncertainty (e.g., what is the object's mass) and dynamic uncertainty (e.g., the object's momentum; the roughness of the surface it moves on). A noiseless simulation with high fidelity fails to capture people's intuitions in physical reasoning tasks [6, 32]. If collision detection is a separate module within the mental physics engine, it likely acts as an independent source of uncertainty. In line with this prediction, recent work suggests collisions independently contribute to the noise in a mental simulation [74].

Body and Shape physics engines have separate data structures for the visual representation of an entity (shape) and the physical representation of that entity (body). The shape of an entity is ultimately rendered and displayed graphically, and it can be made of polygon meshes, subdivision surfaces, and so on. The body holds physical properties such as mass, position and friction, and an approximation to its visual shape for the purposes of calculating dynamics and collision detection. To appreciate the difference between body and shape, think of two rubber ducks colliding (as in Figure 2E). As a graphical representation, the ducks can be captured with high fidelity by means of a polygon mesh and textures, but for the purposes of quickly checking and resolving overlaps, other representations such as convex hulls, bounding boxes or other approximate shapes are more appropriate.

When recognizing and categorizing an object people may call on the more detailed shape representation, but when simulating an object moving forward in time, people might only roughly approximate its shape by using simpler meshes or solids. These separate representations may map onto the separate visual systems proposed by [33, 34], with the vision-for-perception pathway being similar to the shape representation, and the vision-for-action pathway being similar to the body representation. The distinction between bodies and shapes seems particularly illuminating of a set of findings in cognitive development showing that infants below 12 months do not use detailed shape representations to track object identity [e.g. 35, 36].

Box 2. Learning Physics

The mental physics-engine hypothesis is agnostic about how the knowledge captured by this engine is acquired. Are people innately equipped with a physics engine attuned to the dynamics of our three-dimensional roughly Newtonian world, with the right priors on gravity, friction, mass, and so on? While it might be evolutionarily useful, such a fully specified innate model is at odds with developmental findings, showing that infants acquire many basic physical notions during the first years of life [16, 76]. How could something like a mental physics engine be learned over development, and to what extent does the same mechanism continue to support learning of new physical concepts and relations later in life?

It is possible that young children have or acquire early on the most basic categories of a physics engine – that the world is parceled into objects, that the dynamics of objects and their interactions is governed by something like forces – but still lack strong expectations about any of the specifics, such as the existence of certain properties, the shape and structure of the forces, the form of motion constraints, the prior distributions over mass and friction, and so on. Under this view, children's developing knowledge of physics may be driven by becoming more certain about these underlying dynamic variables.

For example, consider infants' learning trajectory regarding support events [16]. Infants seem to initially expect objects with any contact to a supporting base to remain stationary. Infants gradually become more sensitive to whether the contact is at the top of the support, then to the amount of contact, and finally to the shape of the object that determines whether its center of mass is roughly over the supporting base (see Figure II). This trajectory has been explained as the acquisition of decision-rules over perceptual variables (rules such as '*if contact is less than mid-point, predict falling*' [17]). But the same trajectory could be explained as the growth of infants certainty concerning the existence and strength of dynamic variables such as joints that could attach the object and support, random environmental forces, a global force like gravity, and the object's bounding body.
But the same trajectory could be explained as the growth of infants certainty concerning the existence and strength of dynamic

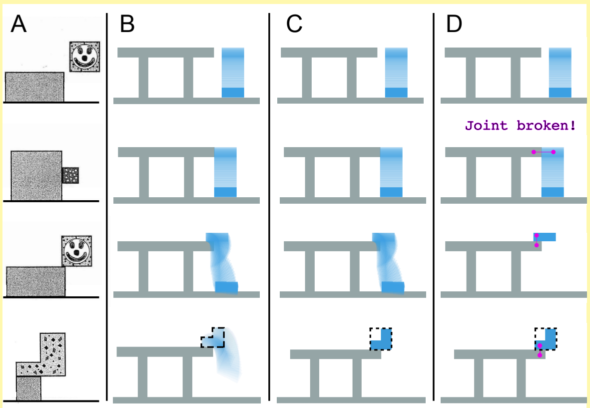

Figure II. Learning About Support. Predicted trajectories for different starting conditions and dynamic assumptions (A) Starting conditions shown to infants, adapted from [16] (B) 'Correct' trajectories expected by infants from about 12.5 months onwards, and by a physics engine with correct assumptions and bounding bodies (dashed) (C) Representing bodies using their bounding box (dashed) leads to the incorrect prediction that L-shaped objects will be supported, as expected by infants younger than 12.5 months (D) The expectation that there may be joints 'gluing' objects together leads to the incorrect prediction that precarious objects on top of a support will stay put, and the correct prediction that objects on the side will fall, in line with the expectations of 5-month old infants.

In the seminal finding, infants see a toy duck and a toy truck appear and disappear, in sequence, from behind the two sides of a single wide occluder. The occluder is then removed to reveal either one or two objects (Figure 2E). One-year-old infants are surprised by the absence of the second object, but younger infants are not. Various controls establish that this failure is not explained by limitations to attention, memory, or general capacities to track objects over occlusion. For example, four-month-old infants are surprised in the above situation if the two distinct objects had moved into view from behind two narrow occluders that were separated by a gap [37]. A great deal of research has elaborated on these original findings, although no single account currently unites all the findings (see [38] for a review of the literature and a physics-based account of it).

When objects such as ducks and trucks are fully visible, infants at 10 months can readily distinguish them perceptually. But for tracking, infants at this age might rely on the body-representation of an object, using similar shape approximations for toy ducks, trucks and other comparable objects (Figure 2E). Such a body-representation proposal is in line with the 'structural layer' proposal [38]. This proposal further predicts that alternating between two shapes with wholly different bodies (such as duck to long spiky snake) or different physical categories (such as rigid body to liquid or soft-body) would lead to different tracking expectations than the duck-truck experiment. In accordance with this last prediction, recent work [39] has shown that young children under memory load notice transitions from rigid to non-rigid states (toy duck to goo) but not similar-shaped rigid transitions (toy car to shoe).

Constraints A physics simulation will often use constraints to restrict the movement of bodies without explicitly calculating forces of motion. Consider a two-bodied pulley system with unequal weights at opposite ends: A physics engine can avoid computing the exact tension on the rope necessary to simulate a force that pulls one mass up while the other goes down. Rather, the engine can enforce a constraint such as 'to the degree that one object moves up, the other moves down'. Common constraints include keeping objects at a particular relative distance (rod constraint), limiting their relative rotation (hinge constraint), constricting objects to move along particular dimensions (planar constraint), or about a particular rotation axis (axle constraints). A common use for constraints is as simple object-to-object attachments, which 'glue' them together. Such 'joints' do not cause the two objects to form a single entity, and the attachment can be broken if the engine detects a threshold of stress or torque has been passed (and see Figure 5C for a hypothetical use of such a joint in explaining infant reasoning about support). Constraints can also be concatenated to create things such as vehicle wheels, pulleys, and chains [12]. Such constraints offer a way of integrating proposals from the field of 'qualitative physics' [40] within quantitative mental simulations.

Hard things, Soft Things, and Stuff Most physics engines classify entities based on their ability to deform, distinguishing between rigid bodies, soft bodies and fluids. Each category is handled differently and requires a varying amount of resources. Fluids and soft bodies are harder to simulate than rigid bodies, and they take up more computation. From early in development, humans also seem to have different expectations about substances compared to objects, and about rigid objects compared to flexible ones [41]. For example, infants do not track piles of sand and flexible compounds in the same way as rigid objects of otherwise similar appearance [42], although they are able to detect changes to the volume of a liquid or non-solid substance [43, 44].

Box 3. Limits of Mental Physics Engines

Early research on intuitive physics suggested that people's reasoning about object motion fails to accord with Newtonian principles, and is subject to surprising errors, even on simple motion prediction tasks [77, 78]. It was later shown that when using more realistic displays and actions, people's intuitions actually closely match Newtonian dynamics [79, 80]. Similarly, earlier work was taken to show that humans use simple heuristics when making mass judgments from dynamic collision events [e.g. 81, 82, 83], but these findings can be subsumed by models based on noisy Newtonian dynamics [10].

Even if the domain in which people can richly simulate physics in their minds turns out to be larger than some have argued, this does not imply that mental simulation is the sole underlying representation for all dynamic reasoning. Some dynamic tasks can be solved quickly through qualitative reasoning without any quantitative simulation [40], and some dynamic tasks – such as those involving wheels and other spinning objects – are difficult for humans to simulate [see for example 54]. Even in inference tasks where physics engines can be useful for evaluating candidate hypotheses or explanations, there remains the difficult and separate problem of coming up with the right hypotheses in the first place [84, 32, 85]. For example, people can reasonably evaluate how well the existence and position of unseen attractors and repellers explain the motion path of objects, but only if they are told this information explicitly. People have more difficulty coming up with the correct hypothesis for the existence and positions of attractors and repellers on their own.

Infants also expect liquids to pour through holes in barriers, and to split and come together, whereas rigid objects should not [45, 46]. Infants extend some of these expectations to non-liquid, non-solid substances such as sand [47]. In many situations, however, infants fail to track non-solid substances over occlusion [42]. Again, physics engines can provide an underlying rationale for why infants find simple tracking of non-rigid objects to be more difficult, because of the resource demands of simulating the movements of liquids and soft bodies.

Physics engines can also provide a computational footing when examining physical concepts within the categories of solids and non-solid substances. For example, developmental researchers have asked whether infants treat sand, water, and honey as entirely separate concepts, as distinct sub-categories of the non-solid substance concept, or as points in a single space of possible non-solid substances varying in their properties [47]? Physics engines can use systems of particles to simulate the behavior of all these non-solid substances, by varying particles' dynamic properties and interaction forces (see Figure 2F), and in this sense they implement a single overarching space of possible non-solid substances. But they can also use different approximations to maximize the efficiency and quality of their simulations for different kinds of substances (and different physics engines can simulate fluids using different approximations); in this sense, they suggest it may be useful to represent distinct sub-types of non-solid substances. Also, many useful expectations about the behaviors of rigid solids, soft solids, and non-solid substances may emerge from a physics engine representation without being explicit. For example, to know that a liquid will form a large puddle when poured from two nearby containers while sand will form two piles (Figure 2F), a physics engine does not need to explicitly contain a 'principle of accumulation' that is specific to the sand concept [48]. Rather, the engine simply needs to run a simulation forward, and examine the result. Such a fluid-simulation may also explain adult proficiency with predicting certain fluid dynamic tasks [49].

By considering physics-engine object classes, we can also propose new mental physics categories to examine in infancy. For example *Cloth*, in the sense of an open mesh that can drape other objects, is particularly difficult to simulate, but abounds in everyday human environments. Cloth is a separate category in most physics engines that are equipped to simulate it, distinct from compressible bodies and fluids. Other entities include fog and smoke, which share certain characteristics with fluids as they can pass through some barriers, compress and split apart, but are not as cohesive as liquids. Similar considerations apply for more one-dimensional entities such as strings, bands, cords, and hair.

Containment Our concept list so far has been one-directional, from physics-engine software to possible mental concepts. However, some categories uncovered by cognitive scientists may be useful for engineers and software developers. As an example, the notion of containment appears relatively early in human development [50, 51]. This category is distinct from visually-similar category of occlusion [52]. In both occlusion and containment events a visible object is visually overtaken by another object. But if the second object is moved, we expect a contained object to go with it, and an occluded object to stay put. Even young infants show these expectations, and seem to further distinguish between loose-fitting containment events and tight-fitting containment events.

If a ball is placed in a box and the box is moved, it may not be worth the computational cost to simulate the ball's motion inside the box. It is sufficient to maintain a simple containment relation, such that the ball's position is linked and updated along with the box's position. Such a work-around can potentially be of use for speeding up physics-engine software.

Physical Illusions

Physical illusions refer to persistent mistaken perceptions in the domain of dynamic reasoning, that clash with people's higher-level belief about the ground truth. Much like visual illusions, physics illusions offer a window into the simplified assumptions made by the computational processes that underlie perception. In particular, it is possible to explain at least some of these illusions by referring to algorithms and assumptions of a physics engine.

As a first example, consider the tall tower shown in the red box of Figure 2B. Most participants agree that this tower is unstable and likely to fall down, while in fact it is stable. Even when people accept as a fact that such formations are stable, they may still 'feel' as though they should collapse imminently. Such intuitions are the basis of an art form known as 'rock balancing' (and see also Figure 3A). These intuitions can be explained by the uncertainty involved in the reconstruction and prediction process of a physics engine [6]. That is, the reconstruction has some degree of uncertainty over the exact position and properties of the objects in the scene. This noise is enough to make the physics engine predict certain stable configuration is in fact unstable, in line with people's intuitions.

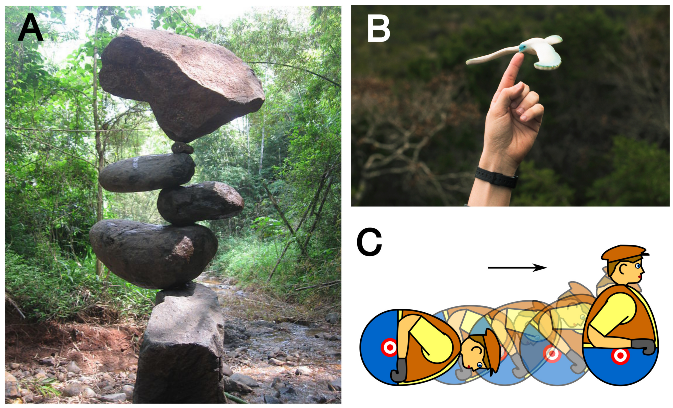

Figure 3: **Examples of Physical illusions.** **(A)** Rock balancing creates precarious-looking stable structures **(B)** Balance-toys are surprisingly supported **(C)** Roly-poly toys seem to lift themselves back up.

Next, consider the stability illusions that underpin popular children's toys, such as the balancing bird shown in Figure 3B and the roly-poly toy in Figure 3C. We expect the bird to tip over, but it stays balanced [53, 6]. We expect the roly-poly to stay tipped over, but it springs back up. When accurately recreated in a physics engine, such objects behave in line with their real world counterparts. However, if we assume that a physics engine creates a simplifying bounding box or convex hull around the shape of a object (see Bodies and Shapes segment above), and makes the simplifying assumption that the density of the box/hull is uniformly distributed, then the objects behave in line with incorrect psychological expectations. For the roly-poly, the center-of-mass is incorrectly located away from the bottom, causing the expectation that it will stay lying down. For the balancing bird, the center of mass is incorrectly located further away from the tip, causing the expectation that it will tip over. Other physics-related illusions discussed as possibly originating from simplifying physics-engine assumptions are the size-weight illusion [11] and the expectation that a wheel rim will roll down an inclined plane at the same speed as a disk [6, 54].

Intuitive Physics in AI, Machine Learning

The need for common-sense reasoning about physical systems as a building-block of intelligence has a long history in AI [see for example 55, 56, 5]. In part, this history stresses the need for defining a dynamic problem in qualitative terms: people know that water put in a heating kettle will boil over time, and that pouring too much water in might cause the kettle to overflow, even if they do not know exactly how and when this boiling and overflow will happen. Similarly, the desired artificial intelligence was to reason over qualitative dynamics and derivatives.

More recently, with the resurgence of artificial neural networks and connectionist architectures across many areas of machine learning [57], there has been a great deal of interest in trying to capture dynamic reasoning with bottom-up approaches that map directly from physical observations to motion prediction or physical judgments. As an example, consider how the Facebook PhysNet architecture tries to capture tower stability judgments [58]. This feedfoward network was provided with many thousands of still images of block towers, which were labeled according to those that did or did not fall under gravity (similar to [6]). PhysNet was able to achieve super-human performance in judging the stability of new towers. This result may be useful for limited AI settings, but it belies that fact that the network does not generalize well even to quite similar scenes (for example, in judging towers composed of more blocks than the training set), nor does it display asymmetries shown by both humans and physics-engine based models [59]. Other networks have been trained to predict the effects of forces from still images [60, 61], and as part of an unsupervised action-guiding predictor of pixel-motion [62, 63] and physical properties [64].

While such networks can achieve success within their domain of training, and may provide a step towards artificial systems with common-sense reasoning, they nevertheless currently lack key aspects of human reasoning that would allow them to generalize flexibly across many different scenarios [65]. Networks such as PhysNet are not reasoning about blocks, mass, friction and gravity; they are reasoning about pixels – abstract patterns in how pixels change over time, but still, pixels. Unlike representations based on explicit objects, relations and events, these image-based representations may not easily extend what has been learned to situations with more blocks, or objects of different sizes and shapes, or the many different inferences human can make, such as predicting which way the blocks will fall, or how many will wind up on the floor, reasoning about which block made another fall over, or understanding how their dynamics might differ if some objects were heavier, smoother or bouncier. This certainly does not mean neural networks have no role to play in intuitive physics. Several groups have recently explored productive ways to combine deep networks with physics-engine based models, such as using physics engines for explicitly simulating the scene's dynamics but vision algorithms based on deep networks as a fast bottom-up initialization of the simulation's state [see for example 66, and Figure 4], or using neural networks to learn the dynamics of forces in a physics-engine-like model that explicitly factorizes into representations of individual objects, their properties and interactions [67, 68, 69].

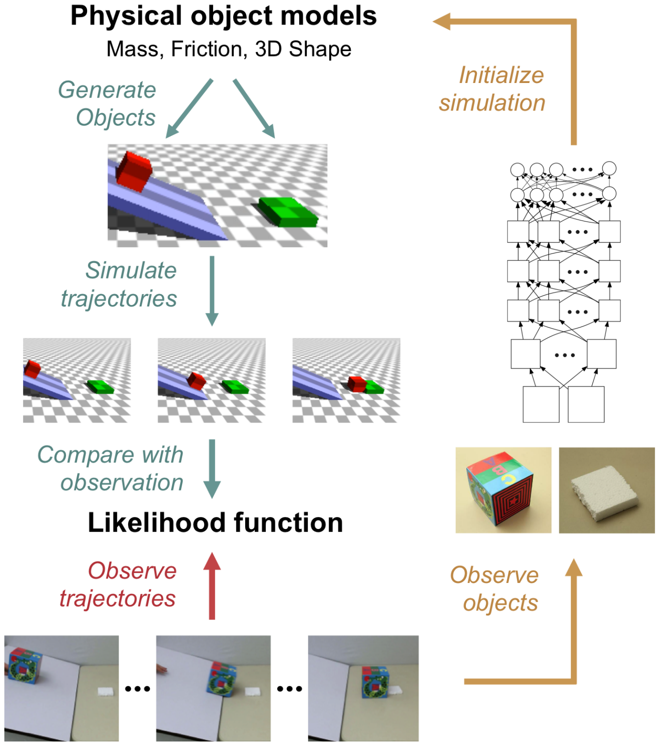

Figure 4: **Inferring Physical Parameters.** From top left: the Galileo system assumes a distribution over physical properties such as mass, and shape. In the forward direction, the system generates objects in space, and simulates their trajectory using a game engine. The simulated trajectory is compared to a real trajectory of objects in motion (bottom left), resulting in a likelihood for the simulated trajectory. The physical parameters are adjusted to maximize this likelihood, and better match observation . In parallel, a neural network is trained to predict the physical properties of objects, given their visual appearance (right). The prediction of the network is used as the initial 'guess' for the physical parameters of the game engine, speeding up inference. For full details see [66].

A Physics Engine in the Brain?

What are the neural substrates of the mental physics engine? Do they form a specific sub-module in cortical processing, or are they part of a broader network? To date, there have been few studies looking directly at the neural signatures of intuitive physical perception and prediction, with research focusing more on the neural representation of explicit textbook physical concepts such as momentum [70], or the brain mechanisms involved in parsing mechanical reasoning puzzles and educational videos of textbook concepts [71]. One recent study has explored the neural basis of more perceptual physical inferences, similar to those used in studies of infants, with a suite of visual scene understanding tasks such as predicting the stability of towers, or predicting the immediate future of simple physical interactions in 2D displays. These tasks were found to preferentially engage a brain network of parietal and premotor regions, apparently overlapping with regions related to action planning and tool use [72]. This finding is in line with previous work showing that visual information about the weight of objects, a key dynamical variable in intuitive physics and game engine simulations, can lead to activation in premotor cortex [73]. An additional experiment in [72] found that the amount of physical content in a video during passive viewing predicts the activation of the brain regions identified as candidate physics-related areas. These results suggest that brain regions relevant for processing intuitive physical inferences are involved in both the perception of scenes and objects, and in action planning and understanding. But these experiments also focused on only a small set of physical inferences, specifically about rigid bodies, and there are still many open questions regarding the neural realization of a mental physics engine.

Concluding Remarks

People do more than classify objects: They see bodies with physical properties, interacting through a play of dynamic forces against a background of inert extended surfaces. Things can be heavy, firm, billowing, fragile, cushy, bouncy. They can fall and smash and blow and drag and flit and anchor. Stuff can ooze and splash and dribble and billow. Because the human mind has to overcome resource challenges when constructing and reconstructing dynamic scenes, we might expect a convergent evolution of concepts between faculties of the mind and simulation software. Taking the mental physics simulation proposal seriously means we should examine the concepts and workarounds that clever people working on game engines develop and use to get their models to work efficiently: concepts whose effectiveness depends both on the nature of the physical world, and on human psychology, but that were developed independently of findings or theories in cognitive psychology. In particular, we should look for those concepts that are shared across many physics engines, regardless of specific implementation details. We examined several such prominent concepts and their design principles, finding new points of inspiration, new perspectives on old phenomena in psychology, and new hypotheses for how intuitive physics might work in the brain and be built into intelligent machines.

Acknowledgments

This work was supported by the Center for Brains, Minds, and Machines (CBMM), funded by National Science Foundation (NSF) Science and Technology Center (STC) award CCF-1231216, and ONR MURI N00014-16-1-2007. The billiard balls in Figure 1A use artwork based on a render by Joey Bennett. The Lego blocks in Figure IA use artwork based on a render by Antonio Stipia. Figure 4A is by Leandro Inocencio (own work) and Figure 4B is by APN MJM (own work) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons.

References

[1] Craik, K. J. W. (1943). *The nature of explanation*. Cambridge University Press.

[2] Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends Cogn. Sci.*, 8, 280-285.

[3] Marcus, G. F. and Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychol. Sci.*, 24, 2351-2360.

[4] Goodman, N. D., Frank, M. C., Griffiths, T. L., Tenenbaum, J. B., Battaglia, P. W., and Hamrick, J. B. (2015). Relevant and robust: A response to Marcus and Davis (2013). *Psychol. Sci.*, 26, 539-541.

[5] Weld, D. S. and De Kleer, J. (2013). *Readings in qualitative reasoning about physical systems*. Morgan Kaufmann

[6] Battaglia, P. W., Hamrick, J. B. and Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci.*, 110, 18327-18332

[7] Gerstenberg, T., Goodman, N. D., Lagnado, D. A. and Tenenbaum, J. B. (2012). Noisy Newtons: Unifying process and dependency accounts of causal attribution. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*

[8] Smith, K. A. and Vul, E. (2013). Sources of Uncertainty in Intuitive Physics.*Top. Cogn. Sci.* 5, 185-199

[9] Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J.B. and Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, 332, 1054-1059

[10] Sanborn, A. N., Mansinghka, V. K., and Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects.*Psychol. Rev.* 120, 411

[11] Hamrick, J. B., Battaglia, P. W., Griffiths, T. L., and Tenenbaum, J. B. (2016). Inferring mass in complex scenes by mental simulation. *Cognition* 157, 61-76

[12] Gregory, J. (2009). *Game engine architecture*. CRC Press

[13] Sutskever, I., Hinton, G. E. and Taylor, G. W. (2009). The recurrent temporal restricted Boltzmann machine. In *Adv. Neural. Inf. Process. Syst. 22*, 1601-1608

[14] Spelke, E. S., Breinlinger, K., Jacobson, K. and Phillips, A. (1993). Gestalt relations and object perception: A developmental study. *Perception* 22, 1483

[15] Spelke, E. S., Breinlinger, K. ,Macomber, J. and Jacobson, K. (1992). Origins of knowledge.*Psychol. Rev.* 99, 605

[16] Baillargeon, R. (2004). Infants physical world. *Curr. Dir. Psychol. Sci.* 13, 89-94

[17] Wang, S. and Baillargeon, R. (2008). Can infants be "taught" to attend to a new physical variable in an event category? The case of height in covering events.textit Cogn. Psychol. 56, 284-326

[18] Lee, S. A. and Spelke, E. S. (2008). Childrens use of geometry for reorientation. *Dev. Sci.*, 11, 743-749

[19] Lee, S. A. and Spelke, E. S. (2010). Two systems of spatial representation underlying navigation. *Exp. Brain Res.* 206, 179-188

[20] Lee, D. N. and Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Percept. Psychophys.* 15, 529-532

[21] Lee, D. N. (1980). Visuo-Motor Coordination in Space-Time. *Adv. Psychol.* 1, 281-295

[22] Michotte, A. (1963). *The perception of causality*. Basic Books

[23] Schlottmann, A. and Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *Q. J. Exp. Psychol.* 44, 321-342

[24] White, P. A. (2006). The causal asymmetry.*Psychol. Rev.* 113, 132-147

[25] Leslie, A. M. and Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25, 265-288

[26] Spelke, E. (1990). Principles of object perception. *Cognit. Sci.* 14, 29-56

[27] Kellman, P. J. and Spelke, E. S. (1983). Perception of partly occluded objects in infancy. textit Cogn. Psychol. 15, 483-524

[28] Valenza, E., Leo, I., Gava, L. and Simion, F.(2006). Perceptual completion in newborn human infants. *Child Dev.* 77, 1810-1821

[29] Baillargeon, R. (1987). Young infants' reasoning about the physical and spatial properties of a hidden object. *Cogn. Dev.* 2(3), 179–200

[30] Kotovsky, L. and Baillargeon, R. (1994). Calibration-based reasoning about collision events in 11-month-old infants. *Cognition*, 51(2), 107–129.

[31] Kotovsky, L. and Baillargeon, R. (2000). Reasoning about collisions involving inert objects in 7.5monthold infants. *Dev. Sci.* 3(3), 344–359.

[32] Ullman, T. D., Stuhlmüller, A., Goodman, N. and Tenenbaum, J. B. (In press). Learning Physical Parameters from Dynamic Scenes. *Cogn. Psychol.*

[33] Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20-25

[34] Milner, A. D. and Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46, 774-785

[35] Xu, F. and Carey, S. (1996). Infants metaphysics: the case of numerical identity.textit Cogn. Psychol. 30, 111-153

[36] Xu, F. (2005). Categories, Kinds, and Object Individuation in Infancy. In *Building Object Categories in Developmental Time*, 63-89

[37] Spelke, E. S., Kestenbaum, R., Simons, D. J. and Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *Br. J. Dev. Psychol.*, 13, 113-142

[38] Baillargeon, R., Stavans, M., Wu, D., Gertner, Y., Setoh, P., Kittredge, A. K. and Bernard, A. (2012). Object individuation and physical reasoning in infancy: An integrative account. *Lang. Learn. Dev.* 8, 4-46

[39] Zosh, J.M. and Feigenson, L. (2012). Memory load affects object individuation in 18-month-old infants. *J. Exp. Child Psychol.* 113, 322-336

[40] Forbus, K. D. (1988). Qualitative Physics: Past, Present, and Future. In *Exploring Artificial Intelligence*, 239-296

[41] Carey, S. (2009). *The Origin of Concepts*. Oxford University Press

[42] Huntley-Fenner, G., Carey, S. and Solimando, A. (2002). Objects are individuals but stuff doesnt count: Perceived rigidity and cohesiveness influence infants representations of small groups of discrete entities. *Cognition* 85, 203-221

[43] Vanmarle, K. and Wynn, K. (2011). Tracking and quantifying objects and non-cohesive substances. *Dev. Sci.* 14, 502-515

[44] Hespos, S. J., Ferry, A. L. and Rips, L. J. (2009). Five-month-old infants have different expectations for solids and liquids. *Psychol. Sci.* 20, 603-611

[45] Hespos, S. J., Dora, B., Rips, L. J. and Christie, S. (2012). Infants Make Quantity Discriminations for Substances. *Child Dev.* 83, 554-567

[46] Rips, L. J. and Hespos, S. J. (2015). Divisions of the Physical World: Concepts of Objects and Substances *Psychol. Bull.* 141(4), 786

[47] Hespos, S. J., Ferry, A. L., Anderson, E. M., Hollenbeck, E. N. and Rips, L. J. (2016). Five-month-old infants have general knowledge of how nonsolid substances behave and interact. *Psychol. Sci.* 27(2), 244-256.

[48] Anderson, E. M., Hespos, S. J. and Rips, L. J. (2017). Five-month-old infants have expectations for how nonsolid substances transform and persist, In *The 2017 Biannual Meeting of the Society for Research in Child Development*

[49] Bates, C., Battaglia, P., Yildirim, I. and Tenenbaum, J. B. (2015). Humans predict liquid dynamics using probabilistic simulation. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*

[50] Hespos, S. J. and Baillargeon, R. (2001). Reasoning about containment events in very young infants. *Cognition* 78, 207-245

[51] Hespos, S. J. and Spelke, E. S. (2004). Conceptual precursors to language. *Nature* 430, 453-456

[52] Hespos, S. J. and Baillargeon, R. (2006). Décalage in infants knowledge about occlusion and containment events: Converging evidence from action tasks. *Cognition* 99, B31-B41

[53] Cholewiak, S. A., Fleming, R. W. and Singh, M. (2013). Visual perception of the physical stability of asymmetric three-dimensional objects. *J. Vis.* 13, 12

[54] Proffitt, D. R., Kaiser, M. K. and Whelan, S. M. (1990). Understanding wheel dynamics. *Cogn. Psychol.* 22, 342-373

[55] Forbus, K. D. (1981). Qualitative reasoning about physical processes. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, 326-330

[56] Kuipers, B. (1986). Qualitative simulation. *Artif. intell.* 29, 289-338

[57] Lecun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature* 521, 436-444

[58] Lerer, A., Gross, S. and Fergus, R. (2016). Learning physical intuition of block towers by example. *arXiv preprint* arXiv:1603.01312

[59] Zhang, R., Wu, J., Zhang, C., Freeman, W. T., and Tenenbaum, J. B. (2016). A comparative evaluation of approximate probabilistic simulation and deep neural networks as accounts of human physical scene understanding. *arXiv preprint* arXiv:1605.01138

[60] Mottaghi, R., Rastegari, M., Gupta, A. and Farhadi, A. (2016). "What happens if..." learning to predict the effect of forces in images. *arXiv preprint* arXiv:1603.05600

[61] Mottaghi, R., Bagherinezhad, H., Rastegari, M. and Farhadi, A. (2015). Newtonian image understanding: Unfolding the dynamics of objects in static images. *arXiv preprint* arXiv:1511.04048

[62] Agrawal, P., Nair, A., Abbeel, P., Malik, J. and Levine, S. (2016). Learning to poke by poking: Experiential learning of intuitive physics. *arXiv preprint* arXiv:1606.07419

[63] Finn, C., Goodfellow, I. and Levine, S. (2016). Unsupervised learning for physical interaction through video prediction. *arXiv preprint* arXiv:1605.07157

[64] Denil, M., Agrawal, P., Kulkarni, T. D., Erez, T., Battaglia, P. and de Freitas, N. (2016). Learning to perform physics experiments via deep reinforcement learning. *arXiv preprint* arXiv:1611.01843

[65] Lake, B. M., Ullman, T. D., Tenenbaum, J. B. and Gershman, S. J. (2016). Building machines that learn and think like people. *Behav. Brain Sci.*, 1–101

[66] Wu, J., Yildirim, I., Lim, J. J., Freeman, B. and Tenenbaum, J.B. (2015). Galileo: Perceiving Physical Object Properties by Integrating a Physics Engine with Deep Learning. In *Adv. Neural. Inf. Process. Syst.* 28, 127-135

[67] Battaglia, P., Pascanu, R., Lai, M., Jimenez Rezende, D. and Koray, K. (2016). Interaction networks for learning about objects, relations and physics. In *Adv. Neural. Inf. Process. Syst.* 29, 4502-4510

[68] Fragkiadaki, K., Agrawal, P., Levine, S. and Malik, J. (2015). Learning visual predictive models of physics for playing billiards. *arXiv preprint* arXiv:1511.0740

[69] Chang, M. B., Ullman, T., Torralba, A. and Tenenbaum, J. B. (2017). A compositional object-based approach to learning physical dynamics. In *ICLR 5*

[70] Mason, R. A. and Just, M. A. (2016). Neural representations of physics concepts. *Psychol. Sci.* 27(6), 904-913.

[71] Jack, A. I., Dawson, A. J., Begany, K. L., Leckie, R. L., Barry, K. P., Ciccia, A. H. and Snyder, A. Z. (2013). fMRI reveals reciprocal inhibition between social and physical cognitive domains. *Neuroimage* 66, 385-401.

[72] Fischer, J., Mikhael, J. G., Tenenbaum, J. B. and Kanwisher, N. (2016). Functional neuroanatomy of intuitive physical inference. *Proc. Natl. Acad. Sci.* 113, E5072-E5081

[73] Loh, M. N., Kirsch, L., Rothwell, J. C., Lemon, R. N. and Davare, M. (2010). Information about the weight of grasped objects from vision and internal models interacts within the primary motor cortex. *J. Neurosci.* 30(20), 6984-6990.

[74] Smith, K. A. and Vul, E. (2013). Sources of uncertainty in intuitive physics. *Top. Cogn. Sci.* 5, 185–199

[75] Smith, K. A. and Vul, E. (2013). Consistent physics underlying ballistic motion prediction. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*

[76] Kim, I.-K. and Spelke, E. S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Dev. Sci.*, 2, 339-362

[77] McCloskey, M., Caramazza, A. and Green, B. (1980). Curvilinear motion in the absence of external forces: naive beliefs about the motion of objects. *Science* 210, 1139-1141

[78] Clement, J. (1982). Students preconceptions in introductory mechanics. *Am. J. Phys.* 50, 66-71

[79] Nusseck, M. and Bardy, B. (1987). Perception and prediction of simple object interactions. *Biol. Cybern.* 27-34

[80] Kaiser, M. K., Proffitt, D. R., Whelan, S. M. and Hecht, H. (1992). Influence of animation on dynamical judgments. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 669

[81] Runeson, S., Juslin, P. and Olsson, H. (2000). Visual perception of dynamic properties: cue heuristics versus direct-perceptual competence. *Psychol. Rev.* 107, 525-555

[82] Gilden, D. L. and Proffitt, D. R. (1989). Understanding collision dynamics. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 372-383

[83] Gilden, D. L. and Proffitt, D. R. (1994). Heuristic judgment of mass ratio in two-body collisions. *Percept. Psychophys.* 56, 708-720

[84] Carroll, C. D. and Kemp, C. (2015). Evaluating the inverse reasoning account of object discovery. *Cognition* 139, 130-153

[85] Ullman, T. D., Goodman, N. D. and Tenenbaum, J. B. (2012). Theory learning as stochastic search in the language of thought. *Cogn. Dev.* 27, 455-480